Virginia

Standards of Learning Assessments

Technical Report

2008-2009 Administration Cycle

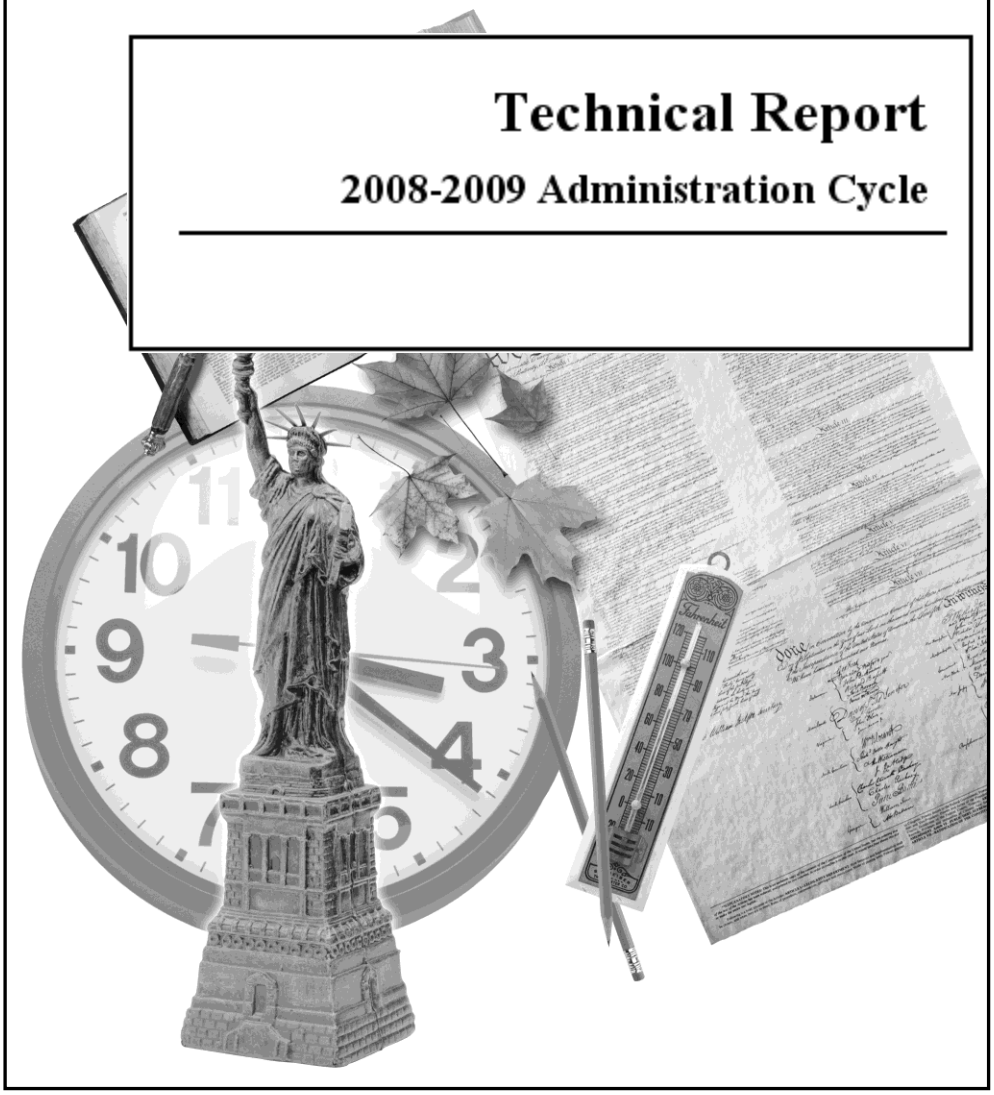


Table of Contents

List of Tables and Figures

PART I: HISTORICAL OVERVIEW AND SUMMARY OF PROGRAMS

1. INTRODUCTION

1.1 Purpose of the Document

The purpose of the *Virginia Assessment Program Technical Report* is to inform users and other interested parties about the development and technical characteristics of the assessments that comprise the Virginia Assessment Program. The *2008-2009 SOL Technical Report* provides information for the 2008-2009 test cycle that comprises the fall 2008 and spring 2009 administrations.

The *2008-2009 SOL Technical Report* is divided into two parts. Part I presents a historical overview of the Virginia Standards of Learning (SOL) and Alternate and Alternative assessment programs and a summary of the components of these programs. Part II is a statistical summary of the 2008-09 administration cycle and an outline of the analyses that were performed. The results for the spring 2009 administration are included here. (Only summary results from new forms are presented here; fall 2008 administration results are based on re-used forms.)

2. STUDENT ASSESSMENTS IN VIRGINIA

2.1 Historical Overview of SOL Assessments

In 1994, Virginia initiated significant reform of its K-12 educational system. This reform, which has evolved over the last ten years, consists of several major elements discussed in the following sections: high academic standards, tests to measure progress, and accountability.

High Academic Standards

In June 1995, after a yearlong development effort, the Virginia Board of Education adopted a set of statewide standards, the Virginia Standards of Learning (SOL). The Virginia SOL set forth minimum learning standards for every child from K-12 in English, mathematics, science, and history/social science. Over time, the SOL were expanded to include the areas of technology, fine arts, foreign language, health and physical education, and driver education.

The Board recognized the need for regular review and evaluation of the SOL, and therefore in September 2000 it approved a cyclical schedule for the review of the standards. This has resulted in each subject area undergoing a review and potential revision every seven years. To date, the history/social science and mathematics Standards of Learning were revised in 2001 and the English and science standards were revised in 2002.

Tests to Measure Student Progress on the SOL

Development of tests to measure the SOL began in 1996 with heavy involvement of classroom teachers, curriculum specialists, and other local educators throughout Virginia. A statewide census field test of the new SOL test items took place in the spring of 1997. The first administration of SOL tests took place in the spring of 1998, and the program has expanded significantly since that time.

The SOL assessment program is the cornerstone of Virginia's system of accountability for the public schools and is authorized in Virginia law and administrative rules (see Article 1, Section 15 of the Constitution of Virginia and Section 22.1-253.13:3C, Code of Virginia). The purposes of the program are to:

- establish and communicate high levels of achievement on the SOL for Virginia public school students;
- provide communication that indicates the progress of students and schools toward meeting achievement levels on the SOL;
- provide information that can be used to improve instructional programs; and
- provide assurance of the quality of public education.

The federally enacted *No Child Left Behind Act of 2001* (NCLB) reinforced many strategies already present in Virginia's public education system. For a number of years, public educators throughout the Commonwealth have focused on instructional standards, student assessment, reporting of results, and continuous improvement. To respond to NCLB, Virginia is maintaining its rigorous academic content standards, measuring students against defined academic performance standards, adding grade-level assessments in various subjects, and reporting Adequate Yearly Progress (AYP) of students at the school, the division, and the state levels. The combination of state and federal educational requirements means that the Virginia Assessment Program will be used to:

a) monitor the progress of students and schools toward meeting established achievement levels;
b) identify educational needs of students;
c) determine which of three achievement levels students have attained (Fails/Does Not Meet the Standards, Proficient in the Standards, Advanced Attainment of Standards);
d) determine whether students receive a high school diploma; and
e) provide accountability information for school, school division, and state levels.

Measures to Ensure Accountability for Student Achievement

In 2006, the Board of Education adopted revisions to the Standards of Accreditation (SOA) for Virginia's public schools. The SOA outlines the requirements for student testing and graduation as well as the requirements for the accreditation of schools in the Commonwealth. The SOA may be found on the Department's website at http://www.doe.virginia.gov/VDOE/Accountability/soa.html.

In 2002, in response to the NCLB, the Board adopted an Accountability Workbook, which outlines the Commonwealth's plan for compliance with the requirements of NCLB and which is updated as amendments to the workbook are approved by the United States Department of Education. The current version of Virginia's Accountability Workbook is available on the Department's website at http://www.doe.virginia.gov/VDOE/nclb/VA-AcctWkbk.pdf

2.2 Overview of Current Virginia SOL Assessments

The Virginia SOL assessments are standards-based tests designed to measure student performance on Virginia's content standards, in the areas of reading, writing, mathematics, science, and history/social science. The SOL tests contain primarily multiple-choice items except for the writing tests administered at grades 5, 8 and high school, which include writing prompts in addition to multiple-choice items.

Addition of Online Testing in Virginia

In the 2000 session of the General Assembly, legislation was passed that required and funded a statewide Web-based Technology Initiative. The goal of this initiative was for Virginia school divisions to implement online, Web-based SOL instruction, remediation, and testing beginning in Virginia's high schools. The initiative provided funding for school divisions to purchase hardware and software and to upgrade network and Internet capabilities.

Because the initial focus of the project was Virginia's high schools, the online testing initiative began with the End-of-Course (EOC) SOL tests. The first online EOC tests were administered in fall 2001. Since that time additional EOC tests have been phased in to the Web-based delivery system so that all EOC tests with the exception of English: Writing are now available in the online system. Virginia's online SOL assessments mirror the paper/pencil SOL assessments in content but are administered to students via a computer. As each SOL test has been implemented in the online system, a comparability study has been conducted to ensure that students are neither advantaged nor disadvantaged by taking the online version of the tests.

Beginning in the 2004-2005 school year, the Grade 8 Science test and the Content- Specific History tests measuring U.S. History to 1877, U.S. History: 1877 to the Present, and Civics & Economics were added to the online delivery system. All middle school SOL tests at grades 6, 7, and 8 with the exception of English: Writing, were available as online tests in 2005-2006, and all elementary school tests were available as online tests in 2006-2007. The volume of online tests administered in Virginia continues to increase with an accompanying decrease in the volume of paper/pencil tests administered.

Current SOL Assessments

The SOL assessments were administered in the 2008-2009 testing cycle to students in elementary and secondary schools. The SOL assessments were administered via the paper/pencil format, and most were also administered online, as noted above.

Students in grades 3 through 8 and high school were tested using multiple-choice SOL assessments in the content areas listed in Table 2.1. In addition, students in grades 5, 8, and high school were tested using a constructed response format in Writing.

Table 2.1 Virginia Standards of Learning Assessments at Each Grade Level

SOL Content Area	Grade Level							
	3	4	5	6	7	8	Content-Specific History	High School
English: Reading	•	•	•	•	•	•		•
English: Writing			•			•		•
Mathematics	•	•	•	•	•	•		
Plain English Mathematics	•	•	•	•	•	•		
History	•							
Science	•		•			•		
Algebra I								•
Plain English Algebra I								•
Geometry								•
Algebra II								•
Virginia and U.S. History								•
World. History I								•
World. History II								•
World Geography								•
Earth Science								•
Biology								•
Chemistry								•
Virginia Studies							•	
U.S. History to 1877							•	
U.S. History: 1877 to Present							•	
Civics and Economics							•	

High school tests were designed to address specific course content, regardless of the student's current enrolled grade. The "Content-Specific History" assessments are not grade-level dependent and are typically taken in the upper elementary or middle school years.

2.3 Content Standards and Assessments

Standards of Learning (SOL)

The SOL represent a broad consensus of what parents, classroom teachers, school administrators, academics, and business and community leaders believe schools should teach and students should learn. In the four core areas of English, mathematics, science, and history/social science,

a curriculum framework is provided that details the specific knowledge and skills students must possess to meet the standards for these subjects.

Curriculum Frameworks

The Standards of Learning Curriculum Framework amplifies the Standards of Learning and defines the content knowledge, skills, and understandings that are measured by the Standards of Learning tests. The Curriculum Framework provides additional guidance to school divisions and their teachers as they develop an instructional program appropriate for their students. It assists teachers as they plan their lessons by identifying essential understandings, defining essential content knowledge, and describing the intellectual skills students need to use. This supplemental framework delineates in greater specificity the minimum content that all teachers should teach and all students should learn.

School Divisions should use the Curriculum Framework as a resource for developing sound curricular and instructional programs. This curriculum framework should not limit the scope of instructional programs. Additional knowledge and skills that can enrich instruction and enhance students' understanding of the content identified in the Standards of Learning should be included as part of quality learning experiences.[1]

3. DEVELOPMENT OF SOL ASSESSMENTS

As noted previously, the Virginia Department of Education works jointly with Virginia educators and its testing contractor to develop a series of tests to measure student achievement against Standards of Learning (SOL) content. The development of the SOL assessments involves the use of test blueprints, item development specifications, multiple review committees, field-testing, and item banking.

3.1 Test Blueprints

The SOL test blueprint serves as a guide for test construction. Each test covers a number of SOL. In the test blueprint, SOL are grouped into categories that address related content or skills. These categories are labeled reporting categories. When the results of the SOL tests are reported, the scores will be presented in terms of scores for each reporting category and a total test score. Each SOL is assigned to only one reporting category.

The number of test items that will be assessed in each reporting category as well as on the test as a whole can be found in the test blueprint. Due to the large number of SOL in each grade level content area, every SOL will not be assessed on every version (form) of an SOL test. By necessity, to keep the length of a test reasonable, each test will sample from the SOL within a reporting category. However, every SOL is eligible for inclusion on each form of an SOL test.

1 The complete curriculum frameworks can be accessed at the following website:

http://www.doe.virginia.gov/VDOE/CurriculumFramework.

The test blueprint also calls attention to any SOL that will be excluded from the test. Some SOL cannot be appropriately assessed in the multiple-choice format.

There is a blueprint for each test (e.g., Grade 3 Reading, Grade 5 Mathematics, Grade 8 Science, U.S. History). Each blueprint contains the three components relevant to each SOL test: development guidelines, a blueprint summary table, and the expanded blueprint. Each of these is discussed in further detail in the following sections.

Test Development Guidelines

Test development guidelines are used by the testing contractor and the members of the Content Review Committees in developing the SOL tests. This section contains three parts:

1. General Considerations — lists general considerations that are used in developing the test as well as considerations specific to a particular content area.
2. Item Format — lists information on how items for the test are constructed.
3. Ancillary Materials — lists any materials (e.g., calculators, rulers, protractors, compasses, dictionaries) that students are allowed to use while taking each test.

Blueprint Summary Table

A summary table of the blueprint displays the following information:

- reporting categories for each test;
- number of test items in each reporting category;
- SOL included in each reporting category;
- SOL which are excluded from the SOL test;
- number of operational items on the test;
- number of field-test items on the test; and
- total number of items (operational and field-test items) on the test.

Expanded Blueprint

The expanded blueprint provides the same information as the blueprint summary table except that the full text of each SOL is included. In addition, SOL that are excluded from the test are categorized by the reason they were not included.

3.2 Multiple-Choice Items

Specifications and Development

Using SOL-specific item specifications, ETS content specialists and contracted item writers construct approximately 300 multiple-choice (MC) items annually for field testing. ETS develops numbers of items for each assessment's reporting categories, generally proportionate to the reporting category percentages in the test blueprints. All items assess content specified by the SOL and within the guidelines contained in the associated Curriculum Framework.[2]

ETS content specialists are responsible for developing MC items that adhere to principles for quality item construction, universal design, and fairness (bias and sensitivity issues). Items are developed for presentation in two modes: online delivery and printed test books. Each item is coded for its SOL. Internal reviews at ETS include at least two rounds of content reviews, a professional editorial review, and a fairness review. Additional guidance and feedback is provided to ETS regarding the appropriateness of the content match to the SOL and adherence to item specifications through Virginia content review committee meetings as well as reviews completed by Virginia Department of Education staff members.

Content Review Committees

On an annual basis, Virginia educators from across the state participate in the development of the SOL assessments. Every summer, Content Review Committees convene in the Richmond area to review content materials for the VA SOL program. Content committees are composed primarily of educators teaching the subject of the test. A small number of committee members may be school test coordinators, curriculum staff, or other school division employees. They represent all grade levels—grade 3 through high school—and content areas, and the racial/ethnic diversity of Virginia students. Committee members also represent a geographical cross-section of Virginia. Every committee has approximately one-third new members introduced each year in order to provide for a balance of experienced educators and new members, and to bring new perspectives into committee meetings. These individuals review the newly developed test items to ensure that they appropriately and fairly measure student knowledge and skills in accordance with the SOL and Curriculum Frameworks.

The committee meetings begin with a general training session conducted by VDOE representatives and an ETS Test Development Manager. Review Committee members receive an orientation to the SOL assessment program, an overview of the test development process, and information about their important role at three major stages – new items, data review of field-tested items, and new test forms. Training focuses on educators making judgments about the match of content to SOL, the appropriateness of the content and difficulty level, and best practices in item construction. VDOE and ETS emphasize the educators' contribution to the

2 Samples of SOL test items are available in released test forms posted on VDOE's website at:

http://www.doe.virginia.gov/VDOE/Assessment/releasedtests.html.

validity of the SOL assessments; they remind committee members that their experience as teachers in the content area is what is valued the most. A significant portion of this general training session is presented via DVD to ensure a standardized presentation.

Following the training session, individual Content Review Committees meet separately by grade level and subject. An ETS content specialist facilitates the committee review, with representatives present from the Department of Education's Office of Assessment Development, as well as Curriculum and Instruction. Prior to reviewing each distinct set of materials—new items, test forms, and item statistics (data review)—a more intense task-specific training is conducted with DVD-delivered presentations, followed by a question/answer session facilitated by the ETS content specialists.

In new item review, items that have not yet been field tested are presented to the committee. The committee reviews them to verify the accuracy of their content, to ensure that the item is aligned to the appropriate SOL and written within the approved specifications, and to ensure the appropriateness of their difficulty level and the quality of item construction and associated graphics/art/stimuli. Content Review Committee members also identify and note their concerns regarding potential item bias in the areas of gender, racial/ethnic, religious, socioeconomic, and regional characteristics. Additionally, special populations concerns may be noted in regard to disabilities and limited English proficiency. Following discussion, the committee as a whole recommends that an item be accepted, edited, or rejected. Each committee member is also provided an individual comment (input) form. While committee decisions are made by consensus, committee members also record their individual recommendation, along with any comments that differ from the committee consensus on the comment forms. All recommendations are tallied, and all comments are compiled into a master document that becomes an official record of the committee review. Only after committee recommendations are counted and comments recorded is the final decision about an item made. As a result of this new item review process, some items are eliminated from the prospective field-test set, while others are edited in the manner directed for field testing.

In addition to the Content Review Committee's bias review, a separate Bias and Sensitivity Review Committee examines each item on the high school tests following field testing. Bias Review committees are convened by subject area (i.e. math, science, English/language arts, and social studies). Committee members are selected from the same pool of applicants as the Content Review Committee, and the same criteria for selection apply.

Before Bias and Sensitivity Review Committees begin to consider items, a video produced with input from Pearson psychometricians provides training in identifying bias in assessments. The training provides examples of cultural, economic, racial, religious, regional, and gender bias. Committee members are challenged to identify instances where bias in an item may affect the performance of an identifiable group of students. In addition, the bias facilitator provides information on how to interpret differential item performance measures to determine if an item is biased against any specific group of students. The committee can also address issues of sensitivity in the test. Any topic or subject that may disturb or upset students, and in the process affect their performance on the test, is a sensitivity issue. The Bias and Sensitivity Review Committee can vote to accept or reject items, and can also recommend addressing problems with bias and sensitivity in future development.

The Special Forms Review Focus Group examines the SOL assessment test forms for students with visual disabilities. Committee members are teachers of visually impaired students. The committee judges the appropriateness of the test format and edits or deletes items deemed inappropriate for students with specific visual disabilities. Based on the decisions of the Special Forms Review Focus Group, Braille and large-print test forms are constructed to accommodate students with visual impairments. Audiocassette tapes of the test forms are also made for students who need them in order to participate in the testing program.

3.3 Writing Prompts

Specification and Development

Students in grades 5 and 8 and high school are tested in writing using a format that requires a response to a prompt. New writing prompts are developed and field-tested to accommodate the operational requirements for new prompts. Generally, writing prompts may be field-tested once in a four- to five-year period. Upon direction by VDOE, ETS English Language Arts content specialists and contracted item writers draft large numbers of potential writing prompts. Writing prompts adhere to SOL specifications and are written in the form of a question, an issue, or a hypothetical situation.

VDOE staff may preview ETS's draft prompts prior to presentation to committee. Input provided to ETS may consist of comments relative to the prompt's clarity, appropriateness for the SOL, non-duplication of prior prompt topics, and perceived ability of the prompt to elicit an extended written student response.

Review Committees

The summer Writing Content Review Committees review newly drafted writing prompts before they are field tested, as well as reviewing data from field-test results for new writing prompts. The review process is similar to that used for review of new items: The committee as a whole provides a consensus recommendation, with individual members' comments captured on a prompt comment form. For new prompts, edits may be made prior to field testing. For prompts with field-test results, the recommendation is to accept or reject each prompt. Committee members determine if the prompts were appropriate for the grade level being tested in terms of difficulty, clarity, general interest to most students, reading level, and perceived ability of the prompt to elicit an extended written student response.

3.4 Field Testing

Once items have been developed by ETS, reviewed and approved by the Content Review Committees and VDOE, they are available for inclusion on a field test. To ensure that sufficient high-quality test items are available for the development of new operational assessments each year, approximately 220 items are field tested annually for each grade and subject. Generally, all

field-test item statistics are captured from spring test administrations, with multiple-choice (MC) items embedded in various core operational forms.

Embedded Design for Multiple-Choice Items, Standalone for Prompts

Field test items are embedded within the test forms in such a way that they appear throughout the operational test form and are not identifiable to students. One operational test form may contain anywhere from one to 18 different sets (versions) of field-test items. In order to field test the number of items needed to replenish the item bank in each grade and subject it is necessary to have these multiple variants of a single operational form.

For the Writing tests, in addition to embedded MC items, writing prompts are field-tested using stand-alone field test administrations, generally occurring in the winter. Pearson psychometric staff develops a sampling plan for the distribution of all multiple-choice test forms with embedded items. The sampling plan is approved by VDOE.

Sampling

During each spring test administration, test forms are distributed throughout the Commonwealth in a way that will facilitate timely equating and the collection of representative field test data. The manner in which test forms are distributed across the school divisions is called the sampling plan. The sampling procedures are based on data files containing participation counts that schools submit to Pearson early in the spring term. These files indicate the number of students in each school who will take each test online or in paper-pencil format. In conjunction with the participation counts, the school division's graduation date and the date that school closes for the year are considered when assigning tests forms in the sampling plan.

An attempt is made to assign test forms to divisions in such a way that approximately equal numbers of students respond to each field test variation across the cores. Also, test forms are assigned at the school division level so that all schools are administered the same core of each test. The core that is assigned to a division by the above process is labeled the "Main" form for that division. Each division is also assigned an alternate form. The alternate form is utilized in retesting students involved in testing irregularities. For instance, an administrator may need to assign a different test form if the student becomes ill during a test or if there is a disruption that prevents the student from completing the test.

The multiple-choice section of the writing tests is assigned to divisions in the same way as the non-writing tests. Two new writing prompts are administered each spring and data must be collected to equate them along with the multiple-choice sections of the test. One of these prompts is designated as the main prompt for the state and the other is designated as the alternate prompt. However, in order to obtain enough data to calibrate the alternate prompt for equating purposes, it is randomly assigned as the main prompt to enough divisions to account for approximately 25% of the students taking a writing test. These divisions are referred to as "Alternate" divisions for that administration of the writing tests.

Data Review Committees

As previously noted, Virginia educators convene each summer to review items. The Data Review Committee members review items with newly generated field-test item statistics from the prior spring administration. Committees receive fairly intensive training delivered via DVD for this task. Additionally, following the DVD, ETS content specialists follow a written script to walk committee members through a review of two or three items, pointing out the meaning and significance of the item statistics. A Pearson psychometrician is available for questions that an ETS content specialist or Test Development Manager is not trained to answer.

The data review training session emphasizes that the committee members' task is to make judgments and recommendations based upon the appropriateness of items' content, using the field-test item statistics to illustrate student performance on the items. Committee members review items again for content validity and adherence to the SOL. While the data is presented with the item, items can only be rejected for content issues. The statistics serve as a guide to possible problems with an item. In data review, committees recommend accepting or rejecting items. Only in a rare circumstance would a committee recommend the need to re-field test an item—with or without an item edit. As with new item review, a comment (input) form is the official record of committee activity.

The same committee that reviews multiple-choice items also reviews writing prompt results. Additional training in this regard includes an introduction to the contents of the data review books. The committee is also familiarized with the scoring rubrics. One of the elements included in the training book is the VA SOL Writing Field Test Prompt Evaluation Form, which is used by Pearson scorers to evaluate the prompts. This form is a hybrid of qualitative and quantitative information that quantifies how scorers think students responded to the prompt. During the scoring process for field-tested prompts, readers and team leaders recorded their observations about the student responses to each prompt. Subsequently, team leaders were responsible for compiling a qualitative report that addressed the following questions:

- Did the students understand what the prompt asked them to do?
- Did the students seem engaged by the prompt?
- Were the students able to effectively focus on a central idea and provide specific information and details?
- Did the readers, based upon reading hundreds of student responses to the prompt, recommend that this prompt be used for live testing?

The report also includes the final score frequency distribution for the prompt. An area for suggestions and comments from the scorers is also included, as are several student responses. In addition, any teacher comments that exist are included in the report as well.

Committee members review the prompt and responses not for content validity and adherence to the SOL, but to ascertain whether the prompt actually elicited responses that are complete and well-elaborated. Members also review the prompt itself for appropriate content and to ensure fairness for all students. A prompt that elicits responses that are similar to lists or a prompt that seems to confuse students is considered to be poorly performing and is usually recommended for

rejection. In some circumstances, a prompt will be recommended for re-field testing at another grade level.

At the conclusion of the meeting, teachers are given an opportunity to evaluate the meeting and record comments about their experience. They are encouraged to provide honest feedback about all aspects of the meeting including process, logistics, and facilitation. These evaluation sheets are compiled and statistics are generated to show the overall satisfaction of the members' experience.

Statistics Reviewed

Descriptive statistics are derived from each spring field test for each test item including classical, Rasch, and differential item functioning (DIF) item statistics for multiple-choice items. Results from the field test administration provide a basis for including items in the operational test forms and constructing equivalent forms.

The statistics calculated from the multiple-choice items included:

- Numbers of students tested;
- Traditional difficulties (*p*-values);
- Item-option response distributions for all respondents by gender and ethnic group; and
- Point-biserial correlations.

Statistics computed for the results of the writing field test included:

- Numbers of students tested; and
- Frequency distributions, means, and standard deviations for the writing domain raw and total scores.

To supplement the traditional statistics, item difficulty parameter estimates based on *Item Response Theory* (IRT) are computed. When using this technique, a statistical model is fitted to the data to estimate item difficulty and item fit.

DIF statistical procedures, such as the Mantel-Haenszel procedure, compute the probability that one demographic group was more likely than another group to answer an item correctly, when the groups are equally able. This information is useful in reviewing items and tests for potential bias. High values of the Mantel-Haenszel Alpha indicate that an item interacts differently among equally able students in the reference and comparison groups. The Mantel-Haenszel procedure compares white and African-American students, white and Hispanic students, and male and female students. The Mantel-Haenszel group differences that exceeds a chi-square significance level of 0.10 are "flagged" for further scrutiny.

A Rasch IRT method of computing DIF statistics provides item difficulty estimates among demographic groups. Under the Rasch model, the only reason for differences in item difficulty statistics is some group characteristic other than achievement. When the Rasch item difficulty estimates between groups are statistically different, further examination is warranted. The Rasch procedure compares white and African-American students, white and Hispanic students, and male and female students. Rasch item difficulty differences exceeding a threshold of 0.52 are

"flagged" for further scrutiny. A detailed description of methods for identifying DIF in test items is in Camilli and Shepard (1994) and Wright and Stone (1979, pp. 192-195). They provide a derivation of the SEM criterion used to flag Rasch item difficulty group differences.

3.5 Item Bank

The SOL item bank is maintained by ETS. The item bank consists of items for all tests coded by SOL (which drives the appropriate test usage). For all passages/items/stimuli, the bank carries the text, art/graphics codes, item codes for required elements (metadata), and historical statistical records for appropriate test administrations. Test items are readily available for ETS's test assembly and for export for Pearson's test form composition.

The metadata information stored for each item in this bank includes:

- unique item identifier (item code);
- SOL test identifier;
- content area;
- grade, Content-Specific History tests, or EOC designation;
- SOL;
- reporting category;
- prior SOL unique item identifiers (VA code, Harcourt code); and
- status (e.g., field-test ready, ready for operational use, released).

Pearson generates statistical records from appropriate test administrations and ETS carries these records in a database in the item bank. Associated statistical records stored for each item in this bank include:

- test form;
- test administration;
- sequence number;
- key;
- reporting category;
- item type (operational, linking, or field-test);
- p-value;
- percentage selecting each distractor;
- percentage omit;
- point-biserial correlation;
- scaled Rasch parameter;
- item fit statistic and flag;
- gender DIF and flags;
- ethnicity DIF and flags; and
- mode (i.e., online or paper).

3.6 Test Construction

Procedures

New core operational test forms are generally used for the first time in the spring administration. Annually, for multiple-choice tests, three core forms are developed for grade 8 reading and math and all EOC assessments. Three core forms are also developed annually for all Writing tests (at grades 5, 8, and EOC). Two core forms are developed for all other SOL tests.

Test specifications are developed by ETS with guidance from Pearson Psychometrics and VDOE. Pearson and ETS jointly develop Test Construction Guidelines for VDOE approval. Test Construction Guidelines provide the operational process and the established expectations (both psychometric and content characteristics) to guide SOL forms assembly. The goal is equivalent test forms within a year and across years.

A common (anchor) item linking design is used year-to-year. One core form from the prior spring administration provides the anchor items placed in the new administration's two or three core forms. Anchor items are placed in the same or nearly-same sequence positions in the new core form, with the exception of the Reading and Writing tests. For Reading and Writing tests, which contain mostly passage-based sets, one anchor passage set is used, and is placed as close as possible to the same sequence as in the prior form. Anchor items represent approximately 30 percent of the operational forms with each SOL test containing from 12 to 18 linking items. ETS content specialists select anchor items, and Pearson Psychometrics and VDOE approve the anchors.

Following approval of anchor items, ETS content specialists select the remaining operational items for each test. During the test construction process Pearson psychometricians ensure that each form meets the test specification blueprint and the statistical targets established in order to ensure equivalent forms within and across years. Quality checks are done to make certain that items on each form have balanced keys and that individual items fall within established guidelines.

These draft forms are reviewed by VDOE and any replacements to items result in a new review by the Pearson psychometricians. This review and edit process continues in an iterative fashion until VDOE has provided final approval.

Review Committees

Once this phase of test development is completed, the newly drafted operational test forms for each SOL assessment are reviewed by the Content Review Committees at the summer meetings. The two or three test core forms are reviewed for each SOL test. Committee members receive training for this task via a DVD presentation. The training focuses committee members on the match to the SOL test blueprint, the arrangement of items within the form, and the balance of topic coverage and item types. Additionally, members are asked to confirm the appropriateness of the item content and accuracy of the keys.

Individual committee members have a comment form to record the numbers of items by reporting categories (to confirm the blueprint match), and to record their overall assessment of the test form as well as comments on individual items. Individual members' comments and recommendations are compiled into a master document following the meeting. Committee review may result in the need for ETS to substitute one or more items.

Any suggestions for changes are subject to review and approval by VDOE. Pearson Psychometrics approves all final item changes to ensure test forms meet psychometric standards. Once operational test cores are final, ETS content specialists embed multiple field-test sets of items into a core form to create the unique versions of the forms that are used for operational administration.

4. TEST ADMINISTRATION

4.1 Training and Materials

To ensure the successful administration of the SOL assessments, VDOE staff provide training to the Division Directors of Testing (DDOTs) before each fall and spring test administration. DDOTs are in turn required to provide appropriate training to the division's School Test Coordinators (STCs). They address training topics for paper/pencil and online tests including procedures for ensuring test materials are kept secure, testing schedules, make-up sessions, and return of test materials.

STCs provide training to the school's Examiners and Proctors based on information made available in the testing manuals, local directions received from the DDOT, and other pertinent sources. They address training topics for paper/pencil and online tests including the review of security requirements, preparation of the test site, and the provision of accommodations for eligible students.

4.2 Testing Windows

For the Fall Non-Writing test administration, each school division administers the SOL tests within its own established "testing window." The DDOT works with the STCs to establish dates and times for each school. For the Spring Non-Writing test administration, each school division administers the SOL tests for grades 3-8 and the content specific tests within one of three statewide established "testing windows." Divisions administering EOC tests may choose their own testing window. The DDOT works with the STCs to identify the dates and times for each school, and the STCs advise the examiners of their schools' test dates and times.

The Writing tests have a designated three-day testing window for each Fall and Spring test administration. During this three-day window, students are administered the multiple-choice and short paper parts of the Writing test. Make-up sessions may occur after the three-day window.

Each school's test schedule has to allow opportunities for make-up sessions to be held prior to the end of the division's testing window.

4.3 Test Security procedures

All persons in the division who have access to or assist with the administration of the paper/pencil or online SOL assessments must read the *Test Security Guidelines*, which outline legislation passed by the Virginia General Assembly[3] regarding the repercussions of violating test security, and sign the test security agreement. This security agreement requires that persons involved in the test administration exercise the necessary precautions to ensure the security of content and all test materials. This agreement must be completed and forwarded to the DDOT before they are given access to the test. These forms are included in each *Examiner's Manual* and the *2008–2009 SOL Assessments Resource Manual.*[4]

Division/school personnel involved with both online and paper testing need to sign only one test security agreement. Persons who have not signed the *School Division Personnel Test Security Agreement (Including Examiners/Proctors)* may not be allowed access to any SOL tests.

WITHOUT EXCEPTION, copies of secure test booklets and writing prompts (including Braille and large-print test booklets, Examiner copies, and audiotapes), and used answer documents **must** be kept in secure, locked storage at all times when they are not in use in an actual testing session.

Each school division must ensure the security of all test materials from the time of receipt until all testing is completed. Secure test materials are shipped and addressed to the DDOT, who has primary responsibility for their security. Information included in all of the SOL manuals explains security procedures pertinent to the receipt, inventory, distribution, and storage of test materials.

4.4 Testing Accommodations

The Individualized Education Program (IEP) Team or 504 Committee has the responsibility for decisions regarding the need for and selection of accommodations for students with disabilities. Similarly, the Limited English Proficient (LEP) committee determines how LEP students will participate in the SOL assessments and what, if any, accommodations are required. Accommodations allow students with disabilities or LEP designation equal access in demonstrating their achievement. Typically, accommodations can be classified in the following categories:

- timing/scheduling;
- setting;
- presentation; and
- response.

Accommodations considered for testing should be those the student uses during classroom instruction and assessments as identified in the student's IEP or 504 Plan, or LEP participation

[3] §22.1–19.1 Actions for violations of test security procedures and §22.1–292.1 Violation of test security procedures: revocation of license.
[4] These manuals may be downloaded from the following website: www.doe.virginia.gov/VDOE/Assessment/home.shtml .

plan. The student should be familiar with an accommodation because the use of an unfamiliar accommodation during testing may have a negative impact on the student's performance. However, it is important to note that certain accommodations used for instruction or classroom assessment may not be allowable on the statewide assessment. Finally, an accommodation based solely on its potential to enhance performance beyond providing equal access is inappropriate.

5. WRITING SCORING

5.1 Staff Involved in Scoring SOL Writing

The constructed response portion of the SOL Writing Assessment is scored by Pearson human readers. Scoring Directors train Scoring Supervisors and Professional Scorers to score the student responses. Highly qualified, experienced readers outside the state of Virginia score all writing samples. These readers are drawn from a database of college graduates who completed the selection process for scorers. The need for ethnic and racial diversity is emphasized throughout the selection process. Scorers for the VA SOL Writing test have a minimum of a bachelor's degree in an appropriate academic discipline (e.g., English, Education), demonstrated ability in performance assessment scoring, and preferably have teaching experience at the elementary or secondary level. The selection process requires that each candidate successfully complete a personal interview, a scoring screening sample, a writing sample exercise, and a grammar test.

Scoring Supervisors are assigned based on proven ability to score responses accurately and communicate scoring standards to Scorers. Scoring Directors are chosen based on their expertise in evaluating writing and their experience training and supervising Scorers. Scorers are trained by Scoring Directors.

5.2 Writing Prompts and How Responses are Scored

The writing samples used for training scorers are from the samples scored during the *range-finding* process. (Rangefinding is the process of identifying model writing samples for the three levels of quality— "domains"—used in the scoring of student writing.) These writing samples, and others identified by Pearson staff and VDOE staff, are used as scoring guides during reader training, qualifying, and calibration. The primary goal of the training is to convey the decisions made during range-finding to the Scorers, and to help them internalize the scoring protocol to effectively apply those decisions.

Training begins with a discussion of the three writing domains used in the scoring model: composing, written expression, and usage/mechanics. The domain-specific training begins with a discussion of the features of the writing domain as well as the score scale. Each response receives a score on a scale of 1-4 points for each of three domains: Composing, Written Expression, and Usage/Mechanics. The four score points represent the following:

- 4 = Consistent control
- 3 = Reasonable control
- 2 = Inconsistent control
- 1 = Little or no control

There are several tools that have been developed to ensure consistent and accurate scoring:

- Reader Bias sheet (This is a document that describes the different ways a scorer could use personal preference instead of assessment standards to score student essays. Scorers are cautioned to avoid bias of any kind.)
- Domains and Definitions sheet (This is a document provided by VDOE that provides descriptions for Composing, Written Expression, and Usage and Mechanics.)
- Scoring Rubrics (one for each domain at each grade level)
- Anchor Sets
- Practice Sets
- Qualifying Sets

Scorers are trained to consider the prompt a "springboard" for the students to write and not to expect any sort of "correct" answer. Scorers also are instructed not to expect any specific mode: If a prompt seems to be asking for a persuasive argument and the student responded with a narrative that was very loosely connected to the prompt, Pearson scores it based on the control of the features of writing, not on how well it responds to the prompt.

5.3 Scorer Training and Qualifying Procedures

Pearson Scoring Directors work with Virginia rangefinding committees to create training sets before each scoring administration. The approved sets are used to train Scorers. As Scorers start training, Scoring Directors review a Reader Bias worksheet developed by Pearson. Scorers are trained to score based on the Virginia rubric's standards and not personal preference. Next, the Scoring Director reviews the Domains and Definitions sheet developed by VDOE. The Scoring Director reviews the various features of the Composing Domain, the Written Expression Domain, and the Usage and Mechanics Domain.

Anchor Sets

For each of the three domains that receive scores, the Scoring Director reviews an anchor set. Anchor sets include 2-4 clear examples of each score point (1-4). The sets are "generic," which means they include student responses to various prompts from many different administrations. Pearson uses these generic sets to ensure consistency from administration to administration. After reviewing each domain-specific anchor set, Scorers practice on 10 sample papers, applying a score for the domain they review. The sets include 2-4 examples of each score point. Scoring Directors provide Scorers with feedback on "true scores" (the scores approved by the range-finding committee) and annotations explaining the correct scores for the practice papers. Scorers then practice on three sets of 10 papers, applying scores to all three domains for each paper. Scoring Directors provide Scorers with feedback based on the scores that they applied versus the "true scores" and annotations explaining the correct score for each practice response.

Qualifying Sets

In order to qualify to score the VA SOL Writing assessment, Scorers take 4 sets of 10 papers and must achieve 70% agreement with rangefinding committee approved scores for each domain on 2 of 4 sets. Scorers who do not meet these standards are released from the project.

5.4 Scoring Procedures

Two Raters with Resolution for Non-Adjacent Scores

In each test administration cycle, all writing responses are reviewed by two Pearson Scorers. The student's score for each domain is the sum of the two Scorers scores. If the two Scorers produced any non-adjacent scores for any domain on a response, then the response is scored a third time by a Scoring Supervisor or Scoring Director. In this case, the student's final score for each domain is determined by the following rules:

1. Calculate the sum of the Third and either the First or Second score, whichever is adjacent to the Third.
2. If both the First and Second scores are adjacent to the Third, use the greater one.
3. If the Third is adjacent to neither, use the Third multiplied by 2.

Validity Checks

Throughout scoring, Scorers receive and score validity papers. These are papers are pre-scored according to rangefinding standards. All scores on validity papers are approved by VDOE. Validity papers are used to monitor consistency in scoring over time; they are interspersed with and indistinguishable from other student responses. VDOE-approved true scores for these papers are loaded into the system, and a report is run that indicates what percentage of accuracy a Scorer achieves on validity papers in scoring against the true score. Validity papers are used as a check to ensure that Scorers, as well as Scoring Supervisors, do not drift from the rubric and continue to score accurately.

Backreading

Backreading is a system which allows a Scoring Supervisor and/or Scoring Director to monitor an individual reader's score. Scoring Supervisors read responses already scored by Scorers they are monitoring. While backreading, the Scoring Supervisor can evaluate the Scorer's performance, provide feedback, and if necessary, override an assigned score. The Scoring Supervisor may also halt scoring activity of an individual or group of Scorers whose performance has declined.

Calibration Sets

Calibration is a process whereby Scorers apply scores to student papers that had been scored previously by a Scoring Director, Scoring Supervisor, and a representative from VDOE. Calibration Sets include 1-3 student responses and are used as a training tool to improve agreement among Scorers. After Scorers take a calibration set, Scoring Directors discuss the correct scores for the responses.

Calibration is a form of training which creates consensus and accuracy within the scoring pool. It is used for the sake of maintaining consistency within a group of Scorers through a mini-training session or discussion based on the given sets. Calibration sets may focus on particular scoring issues including clarifying a scoring line, showing a response that is unusual or problematic to score, or showing a range of responses or performance skill for a particular score point. Scoring Directors present Scorers with a Calibration Set or a review of anchor papers daily throughout scoring.

5.5 Appeals Process

The primary purpose of the appeals process is to provide an additional step to ensure that the score assigned to the student's writing sample produced as part of the Writing tests is an accurate representation of the student's achievement.

Auto Appeals

An automatic appeals process is applied to all Writing tests scored as non-passing that were completed by students attempting to achieve high school graduation by August 31 of that school year. Current criteria of automatic appeals are as follows:

- Student is attempting to achieve high school graduation by August 31 of that school year; and
- A non-passing score was assigned to the student's test; and
- Given the earned score on the multiple-choice component of the Writing test, a passing overall score is attainable when combined with a perfect score on the written component of the Writing test.

Non-Automatic Appeals

Appeals to rescore a student's writing sample may be initiated by parents or by school personnel. All requests for appeals must be reviewed and approved by the school division before being submitted. Requests for appeals should be considered only if there is substantial evidence that the writing sample should have received a higher score. School division staff familiar with the rubric used to score this assessment must review the writing sample. Appeals to rescore such

papers should be approved by the school division only if the reviewers agree that the paper should have received a higher score according to the rubric. A school division may request that a student's writing sample be rescored if:

- the student failed the test, **AND**
- there is evidence that the writing sample produced by the student for the Writing test should have received a higher score. Evidence of this requires that at least two school division staff familiar with the rubric used to score the writing short paper portion of the Writing test review the paper and agree that it should have received a higher score.

6. SCORES AND REPORTS

Following administration and scoring of the SOL assessments, scores are reported by Pearson to students, parents, schools, divisions, and the state. These reports provide a variety of information related to the different types of scores assigned to students.

6.1 Description of Scores

Raw Scores

A raw score represents the number of points a student received for correctly answering questions on a test or for a content area. Because tests may assign different points to questions and have a different total number of questions, the raw score is only useful in relation to that test or content area. For example, consider a student who receives a raw score of 59 on Math and a raw score of 43 on Reading. To put these scores in perspective, there were 75 total items on the Math test and 50 total items on the Reading test. In simple terms, this can mean the respective percent correct would be 79% for the Math test and 86% for the Reading test. Typically, a raw score has meaning only when compared with corresponding scores of a group of students (e.g., a class, school, or age group) or when applied against an accepted criterion or cutoff score.

For the SOL Non-Writing tests that consist of multiple-choice items only, the raw score that an examinee earns is equal to the number of items the examinee answers correctly. For the SOL Writing tests that have a multiple-choice component and an essay component, the raw score of the essay component is calculated as the sum of the ratings given for each element scored,[5] and the total raw score is the sum of the raw scores on the two components (multiple-choice plus essay).

Total Scale Scores

A scale score is a conversion of a student's raw score on a test or a version of the test to a common scale that allows for a numerical comparison between students. Because Virginia uses multiple versions of a test within a grade and subject, the scale is used to control slight variations

[5] Each essay is scored on three elements: composing, written expression, and usage and mechanics.

from one version of a test to the next. Scale scores are particularly useful for comparing test scores over time, such as measuring semester-to-semester and year-to-year growth of individual students or groups of students in a content area. For all SOL tests, the scale scores are set in the range from 0 to 600. A scale score of 0 is set to correspond to a raw score of 0, and a scale score of 600 is set to correspond to a perfect raw score.

Reporting Category Scaled Scores

In order to facilitate the use and interpretation of the SOL assessment results, various scale scores are derived for reporting purposes. Since each assessment covers a number of SOL, the SOL are grouped into categories that address related content or skills in each blueprint. These categories are labeled reporting categories.[6] For each SOL assessment, reporting category scale scores are reported in addition to the overall test scale score. There are varying numbers of reporting categories for the SOL assessments. For each assessment, the reporting category scale scores are set between 0 and 50 with a 30 indicating approximate mastery of the content covered by that reporting category.

Proficiency Levels

In addition to test scores, proficiency levels are reported to individual examinees on all SOL assessments. Examinees are classified into proficiency levels on the basis of their scale scores as compared with the cut scores, which are obtained from the SOL assessment standard setting. For the reading and mathematics assessments in grades 3-8, there are four proficiency levels: Below Basic, Basic, Proficient, and Advanced. For all other SOL assessments, there are three proficiency levels: Basic, Proficient, and Advanced. For all regular SOL assessments the cut score for the Proficient level corresponds to a scale score of 400 and the cut score for the Advanced level corresponds to 500.

[6] A list of the Reporting Categories for a given SOL assessment can be found in the test blueprints, which are located at http://www.doe.virginia.gov/VDOE/Assessment/soltests/

6.2 Reports Provided

Table 6.1 shows the reports provided to students, parents, schools, and districts. The reports are loosely categorized as either student reports or summary reports.

Table 6.1 SOL Assessment Reports Provided

Student Reports	Summary Reports
Student Data Extract by Division	SPBQ[7] Report by School
Student Data Extract by School	SPBQ Report by School by Division
Report to Parents by School	SPBQ Summary Report by School
Report to Parents by Group	SPBQ Summary Report by Division
Report to Parents by School by Division	SPBQ Summary Report by School by Division
Student Performance Report by School	SPBQ Report Summary Record Extract by School
Student Performance Report by School by Division	SPBQ Report Summary Record Extract by Division
Student Performance Report by School	Summary Record Extract by Division
SPBQ Preliminary Report by School	Summary Record Extract by School
SPBQ Extract by School	Summary Report by School
SPBQ Extract by School by Division	Summary Report by Division
On Demand Group List Report	Summary Report by School by Division
	Analysis of Sub-Group Performance Report by School
	Analysis of Sub-Group Performance Report by Division
	Analysis of Sub-Group Performance Report by School by Division

6.3 Appropriate Use of Scores

Raw scores are affected by test length and difficulty. They cannot be used for comparing examinees across different tests or test forms. Raw scores are comparable only within a given test form. While the scale scores can be used for comparisons across test forms within an SOL test, they cannot be compared across different SOL tests. For example, scale scores cannot be used to reliably determine whether a student or group of students is stronger in reading than in mathematics. In the same sense, reporting category scale scores only allow comparisons within a given reporting category.

[7] Student Performance by Question

6.4 Cautions for Score Use

As previously mentioned, for all SOL assessments, the scale scores are constructed so that a score of 400 represents the Proficient cut and a score of 500 represents the Advanced cut. By this means, a standards-referenced interpretation is incorporated into the scale scores. In other words, regardless of what form or administration year of the SOL assessment a student takes, the same level of ability is required to obtain a scale score of 400 for Proficient, and a scale score of 500 for Advanced. For each SOL assessment, the cut scores remain the same over years, but they may correspond to different raw scores across test forms and administrations. The fluctuation of raw scores does not mean that the requirements for the proficiency levels have changed. It only reflects changes of difficulty across test forms and administrations.

7. PERFORMANCE STANDARDS

7.1 Performance Level Descriptors

Performance level descriptors (PLDs) are statements of what a student should know and be able to do at each performance level given the content standards being assessed. In grades 3-8 reading and mathematics, there are four performance levels that a student may achieve: Advanced, Proficient, Basic, and Below Basic. For the EOC assessments, three performance levels exist: Advanced, Proficient, and Fail/Basic.

7.2 Standard Setting Process/Methods and Procedures

Standard setting is defined as a systematic way of making a professional judgment about how many points a student must earn in order to meet a specified criterion, such as the achievement level of "pass proficient."

During the 2008-2009 administration cycle, there were no standard-setting events held.

8. CALIBRATION, EQUATING, AND SCALING

8.1 Calibration Procedures

Item Response Theory

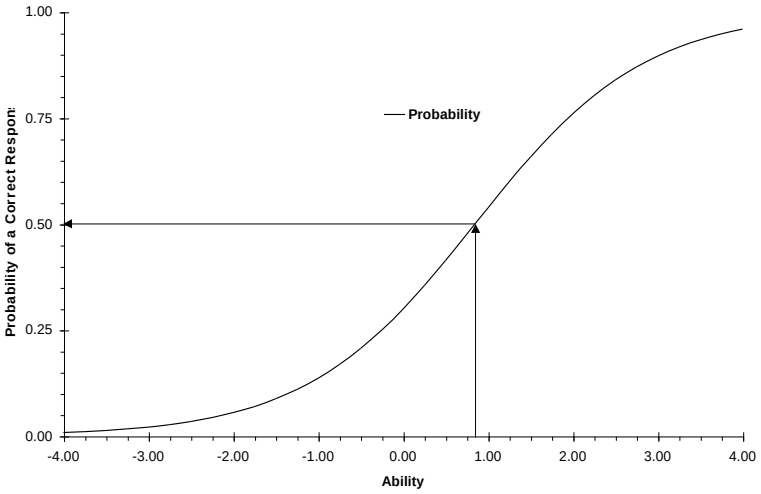
The Item Response Theory (IRT) model used to develop, calibrate, equate, and scale the Virginia Standards of Learning (SOL) was the Rasch model (Rasch, 1980) and its polytomous extension, the Partial Credit model (PCM) (Masters, 1982). These measurement models are regularly used to construct test forms, for scaling and equating, and to develop and maintain large item banks. All test analyses, including item-fit analysis, scaling, equating, diagnosis, and performance prediction, were accomplished within this framework. The statistical software used to calibrate and scale the tests was WINSTEPS Version 3.60.0 (Linacre, 2006).

Rasch and Partial Credit IRT Models

The most basic expression of the Rasch model is in the Item Characteristic Curve (ICC). It shows the probability of a correct response to an item as a function of the ability level. The probability of a correct response is bounded by 1 (certainty of a correct response) and 0 (certainty of an incorrect response). The ability scale is, in theory, unbounded. In practice, the ability scale ranges from -4 to +4 logits for heterogeneous ability groups.

As an example, consider Figure 8.1, which depicts an item that falls at approximately 0.85 on the ability (horizontal) scale.

Figure 8.1 Sample Item Characteristic Curve



When a person answers an item at the same level as his or her ability, then that person has a roughly 50% probability of answering the item correctly. In other words, out of a group of 100 people who each have an ability of 0.85, about 50% could be expected to answer the item correctly. A person whose ability is above 0.85 would have a higher probability of answering the item correctly, while a person whose ability is below 0.85 would have a lower probability of answering the item correctly. This makes intuitive sense and is the basic formulation of Rasch measurement for test items having only two possible categories (i.e., wrong or right).

Figure 8.2 Category Response Curves for a One-Step Item

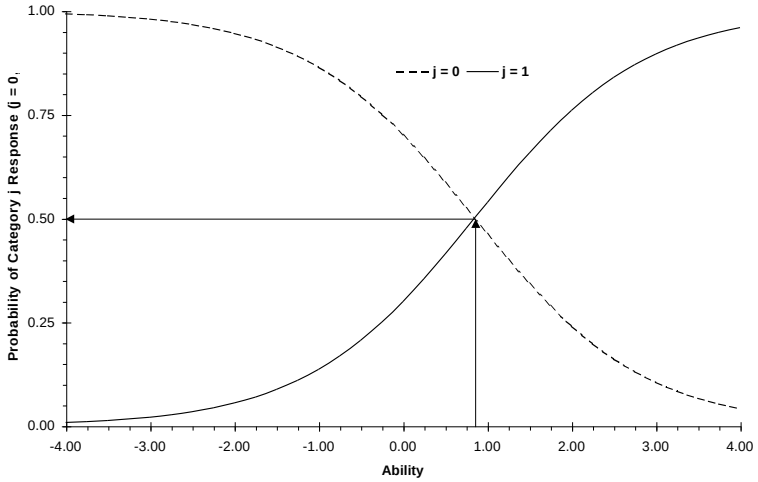
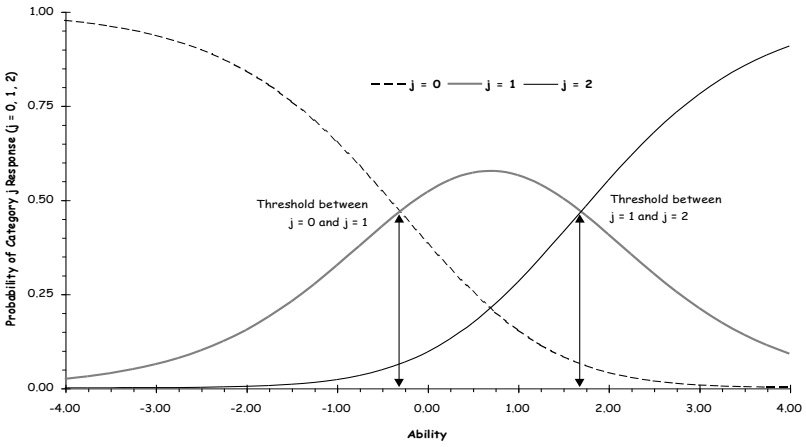


Figure 8.2 extends this formulation to show the probabilities of obtaining a wrong answer or a right answer. The curve on the left (j=0) shows the probability of getting a score of '0' while the curve on the right (j=1) shows the probability of getting a score of '1.' The point at which the two curves cross indicates the transition point on the ability scale where the most likely response changes from a '0' to a '1.' Here, the probability of answering the item correctly is 50%.

The key step in the formulation and the point at which the Rasch dichotomous model merges with the PCM requires us to assume an additional response category. Suppose that, rather than scoring items as completely wrong or completely right, we add a category representing answers that, though not totally correct, are still clearly not totally incorrect. These relationships are shown in Figure 8.3.

The left-most curve (j=0) in Figure 8.3 represents the probability of all examinees getting a score of '0' (completely incorrect) on the item, given their ability. Those of very low ability (i.e., below -2) are very likely to be in this category and, in fact, are more likely to be in this category than in the other two categories. Those receiving a '1' (partial credit) tend to fall in the middle range of abilities (the middle curve, j=1). The final, right-most curve (j=2) represents the probability for those receiving scores of '2' (completely correct). Very high-ability people are clearly more likely to be in this category than in any other, but there are still some of middle and low ability that can get full credit for the item.

Figure 8.3 Category Response Curves for a Two-Step Item



Although the actual computations are quite complex, the points at which lines cross each other have a similar interpretation here as in the dichotomous case. Consider the point at which the j=0 line crosses the j=1 line, indicated by the left arrow. For abilities to the left of (or less than) this point, the probability is greatest for a '0' response. To the right of (or above) this point, and up to the point at which the j=1 and j=2 lines cross (marked by the right arrow), the most likely response is a '1.' For abilities to the right of this point, the most likely response is a '2.' The probability of scoring a '1' response (j=1) declines in both directions as ability decreases to the low extreme and increases to the high extreme.

An important implication of the formulation can be summarized as follows: If the commonly used Rasch model applied to dichotomously (right/wrong) scored items can be thought of as simply a special case of the PCM, then the act of scaling multiple-choice items together with polytomous items, whether they have three or more response categories, is a straightforward process of applying the measurement model. The quality of the scaling then can be assessed in terms of known procedures.

One important property of the PCM is its ability to separate the estimation of item/task parameters from the person parameters. With the PCM, as with the Rasch model, the total score given by the sum of the categories in which a person responds is a sufficient statistic for estimating person ability (i.e., no additional information need be estimated). The total number of responses across examinees in a particular category is a sufficient statistic for estimating the step difficulty for that category. Thus with PCM, the same total score will yield the same ability estimate for different examinees.

The PCM is a direct extension of the dichotomous one-parameter IRT model developed by Rasch in the 1950s (Rasch, 1980). For an item/task involving m_i score categories, one general expression for the probability of scoring x on item/task i is given by

$$P_{xi} = \exp\sum_{j=0}^{x}(\theta - D_{ij}) \Big/ \sum_{k=0}^{m_i}\left[\exp\sum_{j=0}^{k}(\theta - D_{ij})\right] \qquad \text{(Equation 8.1)}$$

in which $x = 0, 1, ..., m_i$, and by definition,

$$\sum_{j=0}^{0} (\theta - D_{ij}) = 0$$

The above equation gives the probability of scoring x on the i^{th} test item as a function of ability (θ) and the difficulty of the m_i steps of the task (Masters, 1982).

According to this model, the probability of an examinee scoring in a particular category (step) is the sum of the logit (log-odds) differences between θ and D_{ij} of all the completed steps, divided by the sum of the differences of all the steps of a task. Thissen and Steinberg (1986) refer to this model as a divide-by-total model. The parameters estimated by this model are (1) an ability estimate for each person (or ability estimate at each raw score level) and (2) m_i threshold (difficulty) estimates for each task with $m_i + 1$ score categories.

8.2 Equating and Scaling Procedures

Rationale

Equating of operational test forms involves ensuring that all forms in a content area and grade level test (e.g., grade 3 Mathematics) are as equally difficult as possible, both within and across assessment administrations. Equating makes certain that students taking one form of a test were neither advantaged nor disadvantaged when compared to students taking a different form of a test.

Common items on each form of the test were used to equate the SOL assessments. Each test form contained a subset of items that was reproduced on every other test form for the same subject and grade. These items, called linking items, served as anchors for comparison. Each time a new test form is constructed in the future, use of linking items ensures the new form will be equal in difficulty to the previous form. Statistical procedures using data collected on items during field tests were used to perform the equating. The data collection design used was the Design IV procedure for common item, non-equivalent groups (Angoff, 1971).

In order to obtain parameter estimates for both the unique items on each form and the linking items, the Rasch model (or PCM) was applied to each test form at a grade level and content area. The parameter estimates for each form were placed on a common metric by using the equating constant procedure (Wright & Stone, 1979). This resulted in the item parameters for all forms being on the same ability scale. A consequence of this was that, given an ability estimate θ, it was possible to determine scores on different forms that could be considered equivalent. The final step consisted of obtaining for each raw score point on a form the ability score or theta corresponding to it. This was done by iteratively solving the expression:

$$TrueScore = \sum_{i=1}^{I} \sum_{j=0}^{m_i} j \cdot P_{ij}(\theta) \qquad \text{(Equation 8.2)}$$

where $P_{ij}(\theta)$ is the probability of a correct response for each of the i=1, ..., I items given that the

item categories are scored *0, ..., m_i*.

Figure 8.4 True Score Equating

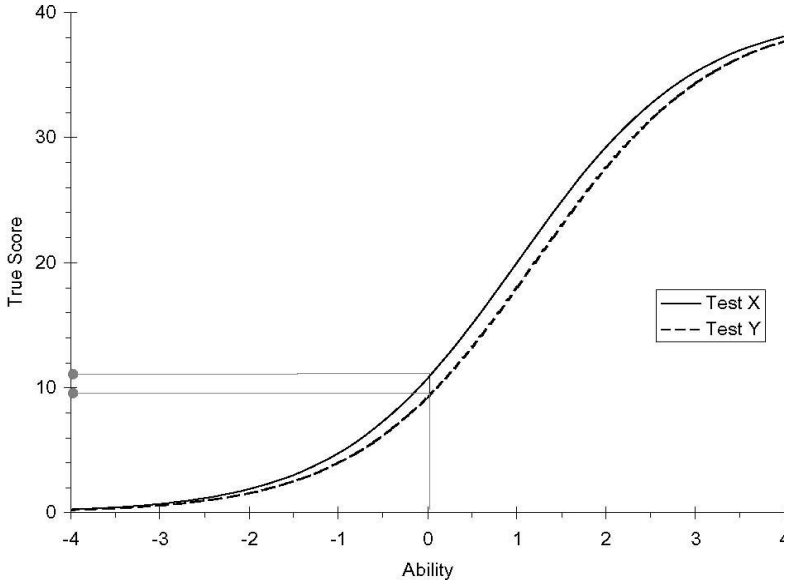


Figure 8.4 illustrates these concepts for two hypothetical test forms, X and Y. In this figure, the true scores on each of the forms are plotted against ability using Equation 8.5. By drawing a line from the ability (here shown for an ability of 0) to each of the respective curves, and moving across to the true score scale, one can find the pairs of true scores that are equated to one another. According to Lord and Wingersky (1983), the procedure applied to true scores can be transferred to observed scores without any major anomalies in the resulting outcomes.

All post-equating on live test forms was carried out at the total score level, and at the reporting category level. Consequently, as new test forms are developed, they will be of approximate equal difficulty at the reporting category level. Members of the Content Review Committees also used data from these analyses for data review.

8.3 Scale Scores

Total Scale Scores for Content Areas

To accomplish the transformation, two levels, d_1 and d_2, were selected on the ability scale corresponding to standards-referenced criteria. These values were converted to the new scale at easy-to-remember locations, D_1 and D_2. Specifically, $D_1 = 400$ was linked to the cut point between *Below Proficient* and *Proficient*, and $D_2 = 500$ was linked with the cut scores between *Proficient* and *Advanced.* Since d_1 and d_2 were criterion values on the ability scale, and D_1 and D_2 were the values on the new scale, the linear transformation (see Wright & Stone, 1979) was given by:

$$ScaleScore = \alpha + \gamma \cdot Theta \quad \text{(Equation 8.3)}$$

where the intercept of the linear transformation is

$$\alpha = \left(D_1 d_2 - D_2 d_1\right) / \left(d_2 - d_1\right) \quad \text{(Equation 8.4)}$$

and the slope is

$$\gamma = \left(D_2 - D_1\right) / \left(d_2 - d_1\right). \quad \text{(Equation 8.5)}$$

This transformation preserved the standards-referenced interpretation of the scale scores by being explicitly linked to the standards-referenced cut scores obtained from the Virginia SOL assessment standard setting. In other words, regardless of what form or administration year of the SOL assessment a student takes, a student would require the same level of ability to obtain a scale score of 400 for proficiency and a scale score of 500 for advanced. While the scale scores can be used for comparisons *within* an SOL assessment, they cannot be compared *across* different SOL assessment content areas.

The scale scores represent a non-linear transformation of the raw scores from which they were obtained. That is, the distance between scale scores does not remain the same for each change in the raw scores. Typically, for the middle of the scale (around the 350 to 400 range), the increments are smaller than near the top or bottom of the scale. To complete the scale, a scale score of 0 was set to correspond to a raw score of 0, and a scale score of 600 was set to correspond to a perfect raw score.

Scale Scores for Reporting Categories

In order to facilitate the use and interpretation of the SOL assessment results, various scale scores were derived for reporting purposes. Since each assessment covers a number of SOLs, the SOLs were grouped into categories that address related content or skills in each test blueprint. These categories are labeled *Reporting Categories*.

Reporting Category scale scores are calculated to provide an interpretation of student performance in each Reporting Category in relation to the performance standard on the test as a whole. The Rasch item difficulty parameters from the full test calibration corresponding to the items comprising each reporting category were used to derive a raw score-to-theta table for each reporting category. Once the raw score-to-theta scale is produced, it is necessary to convert the theta values to scaled scores. The formula used to make this conversion is shown below.

$$\text{Reporting category scaled score} = \left[30 + \left(\frac{\theta_{raw} - \theta_p}{\sigma_{rc}}\right) * 7\right] \quad \text{(Equation 8.6)}$$

Where θ_{raw} is the theta value associated with the reporting category raw score, θ_p is the theta associated with the passing cut on the overall test, and σ_{rc} is the standard deviation of the thetas

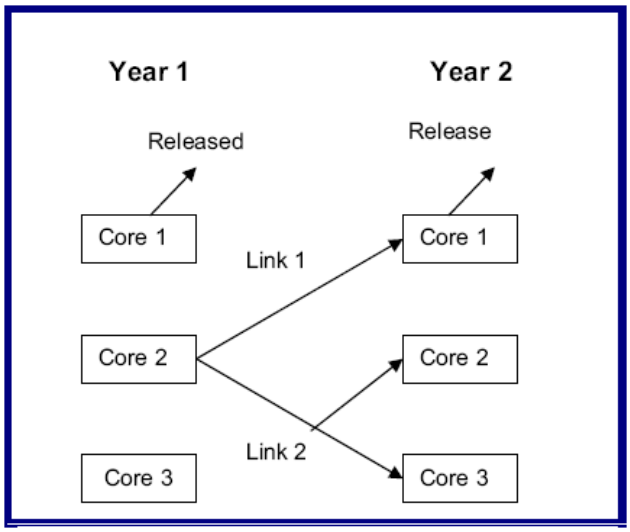
associated with the reporting category. The standard deviations used in these calculations are derived from the first administration of a test after new performance standards are set.

Steps in the Equating and Scaling Process

The equating design is the common-item nonequivalent groups design (Kolen & Brennan, 2005). Under this design, sets of common items called linking items or anchors are placed on two different test forms and used to create the statistical adjustment in scores that equates the first form to the second form. For the Virginia SOL assessments, the number of anchor items in a link is approximately 30 percent of the total items on the assessment.

Figure 8.5 shows the linking process currently used for the SOL assessments. Starting with a set of core forms that are on the SOL score scale, two sets of anchors are used to equate the tests in a subsequent year (i.e., Year 2) to a previous year (Year 1). The Core 2 form in the previous year contributes two sets of anchors. The first set is used to link Core 1 in Year 2 back to Core 2 in Year 1. Under the current linking design, this anchor set is released along with the rest of the Core 1 test to the public, and is thus lost to the program for future use. The second set of anchor items is used to link both Core 2 and Core 3 in Year 2 back to Core 2 in Year 1. Since all three cores in Year 2 are linked to Core 2 in Year 1, this linking across test administrations also achieves the goal of linking alternate versions or core forms to the same score scale.

Figure 8.5 Linking Process for the SOL Assessments



The writing tests have a multiple choice (MC) component and an essay component. Three different cores of MC items are built for each spring administration and each core is paired individually with each of two new essay prompts. Specifically, each core of MC items is paired with each essay prompt, producing six different MC-essay combinations. These are referred to as the Main and Alternate forms. Each Core of MC items contains a set of internal anchor items to link it back to a previously administered writing test. The two essay prompts are selected to have as similar a response distribution as possible based on field test data. A two step process is used to calibrate and equate each of the MC forms and prompts. First, each core of MC items are calibrated and equated separately without including any essay scores using the item parameters from the previous administration to anchor the MC items designated as links. Second, the prompts are calibrated using all students who responded to each, regardless of the MC core they took. The prompt calibrations use the MC item parameter estimates for all three MC cores from the first step as anchors. This requires constructing a separate incomplete data matrix with the MC responses from all three cores and the responses to each prompt. The procedure results in

one set of items parameters for each MC item and prompt. Combining the MC and prompt item parameter estimates for each MC-prompt combination produces a raw score to scale score table for that MC-prompt combination.

The new MC items are also paired with previously administered prompts and previously administered and equated sets of MC items are paired with either new or previously administered prompts. These different combinations are to produce forms for large print and Braille versions of the writing tests, for use in the event that any schools are not able to test on the day that the main essay prompt is administered due to inclement weather, or for the term graduate students at the EOC level. These MC-prompt combinations require only the calculation of a raw score to scale score table, which are derived by pairing the appropriate MC and essay item parameters from the previous administrations or from the current year calibration of the Main and Alternate forms. The MC or the essay items from administrations prior to Spring 2008 have more than one set of item parameters because of the different calibration procedure used. In such a case, the MC or the essay item parameters from the calibration with the largest N-count are used.

The equating of field test items proceeds along somewhat different lines. For this equating, the various alternate forms that contain field test items are linked together by the entire set of core items which perform as anchors during the calibration process to put the field test items on the same scale as the operational items. Because of this, the actual equating of the field test items must take place after the equating of the core forms.

Preparation

Pearson psychometricians prepare the following information for equating and scaling in each administration:

- a description of the core/prompt combinations and form designations;
- test maps which define the content of the SOL assessments, provide the answer keys, and show the relation of items to clusters (i.e., reporting categories);
- WINSTEPS control files generated from Test Map information;
- linking items from the previous administrations;
- cuts scores for SOL achievement levels on the SOL Rasch theta metric; and
- the standard deviations of theta used for reporting category score calculations.

The psychometric team verifies that the anchor items are accurate and prepares files to contain the item parameter difficulties and position information of the linking items. These files are used in subsequent equating steps to anchor the equating calibrations. All WINSTEPS control files are examined and compared to the Test Maps to check for discrepancies. The psychometric staff applies the data exclusion rules that are used for the equating and verifies that the rules have been properly coded in their analysis programs.

Data Checking

Once scores are available for at least 3000 students on an individual core form, an assessment of the data begins. The Pearson psychometric team runs a series of SAS programs to check on the integrity of the data and apply exclusion rules to the data extract. These quality assurance checks include:

- an examination of the raw response frequencies for each item to check for multiple marks and omit rates (multiple choice) and to see if the item shows the proper range of scores (open-ended); and
- a calculation of the *p*-values and point biserial correlations.

Very low (<0.20) or very high (>0.95) *p*-values and very low point biserial correlations (<0.20) are flagged for further scrutiny with content development specialists. For example, a negative point biserial correlation almost always means that an item has an improperly keyed correct answer. Once all content-related flags are resolved by ETS content specialists, equating may proceed.

Raw Score-to-Scale Score Table Production

The psychometric team uses WINSTEPS to calibrate the data. These WINSTEPS runs are anchored using the files of linking items parameters created earlier during the preparation phase. Again, item parameters are estimated and a report is produced that shows for all anchor items their displacement from the item parameter estimates of the previous year. Anchor items with displacements greater than an absolute value of 0.5 are identified as possible outliers and removed from the linking set. The item parameters are then estimated again (without the dropped anchor item) and displacement for the remaining anchor items is evaluated in an iterative fashion until all anchors items have an absolute displacement value less than 0.5. Once the final set of anchor items is decided upon, a final calibration occurs and the resulting output is used to produce the raw score-to-scale score (RSSS) tables needed via scaling efforts (described previously) to provide scale scores for the SOL assessments and associated score reports.

The entire process described above is repeated for each grade and content area, as well as for each reporting category within an SOL assessment. All steps are independently replicated by at least two Pearson psychometricians and an external third party to make certain that there are no errors in calibrating, equating, scaling, or formatting of the RSSS tables.

9. RELIABILITY, CLASSICAL TEST THEORY

There are useful indices available within the framework of Classical Test Theory (CTT) for estimating the precision of the raw test scores and the reliability of assessments. Within CTT, an observed test score is defined as the sum of a student's true score and error ($X = T + E$, where X = the observed score, T = the true score, and E = error). A true score is considered the student's true standing on the measure, while the error score reflects a random error component. Thus, error is the discrepancy between a student's observed and true score.

The reliability coefficient of a measure is the proportion of variance in observed scores accounted for by the variance in true scores. The coefficient can be interpreted as the degree to which scores remain consistent over parallel forms of an assessment (Ferguson & Takane, 1989; Crocker & Algina, 1986). There are several methods for estimating reliability; however, because the Virginia SOL is a secure test that should not be administered twice, an internal consistency method is used. In this method a single form is administered to the same group of subjects to determine whether examinees respond consistently across the items within a test.

9.1 Alpha and Stratified Alpha

The Internal Consistency Method investigates the stability of scores from one sample of content to another by estimating how consistently individuals respond to items. A basic estimate of internal consistency reliability is *Cronbach's Coefficient Alpha* statistic (Cronbach, 1951). Coefficient alpha is equivalent to the average split-half correlation based on all possible divisions of a test into two halves. Coefficient alpha can be used on any combinations of dichotomous (two score values) and polytomous (two or more score values) test items and is computed using the following formula:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{j=1}^{n} S_j^2}{S_X^2}\right), \qquad \text{(Equation 9.1)}$$

where n is the number of items,

S_j^2 is the variance of students' scores on item *j*, and

S_X^2 is the variance of the total-test scores.

Cronbach's alpha ranges in value from 0.0 to 1.0, where higher values indicate a greater proportion of observed score variance is true score variance. Two factors affect estimates of internal consistency: test length and homogeneity of items. The longer the test, the more observed score variance is likely to be true score variance. The more similar the items, the more likely examinees will respond consistently across items within the test.

For the Writing tests, where there is a combination of multiple-choice items paired with an open-ended writing prompt (less homogeneity), a stratified alpha statistic is used to assess the reliability of the assessment. For each item type, the *Stratified Cronbach's Alpha* is calculated using the following formula:

$$\textit{Stratified } \alpha = 1 - \frac{\sum \sigma_i^2 (1 - \rho_{ii'})}{\sigma_t^2} \qquad \text{(Equation 9.2)}$$

where σ_i^2 = variance of scores on item type *i*,

σ_t^2 = variance of total scores, and

$\rho_{ii'}$ = reliability coefficient of scores on item type *i*.

9.2 Standard Error of Measurement

Classical Standard Error of Measurement

The purpose of a reliability coefficient is to estimate the proportion of observed score variance that is true score variance. With this statistic, one can infer the proportion of observed score variance that is error variance. The Standard Error of Measurement (SEM) is another way of understanding reliability. The SEM is the square root of the error variance. This statistic indicates the amount of measurement error in a set of observed test scores. The SEM is inversely related to the reliability of a test; therefore, the greater the reliability, the lower the SEM. With a lower SEM, there is more confidence in the accuracy, or precision, of the observed test scores. The SEM is calculated using the following equation:

$$SEM = \sigma_X \sqrt{1-\rho_{XX}}\ , \qquad \text{(Equation 9.3)}$$

where σ_X is the population standard deviation of observed scores and

ρ_{XX} is the population reliability coefficient.

For a sample of examinees, an estimate of the SEM, when the reliability coefficient is estimated via Coefficient Alpha, is

$$Est\left(SEM\right) = S_X \sqrt{1-\alpha}\ , \qquad \text{(Equation 9.4)}$$

where S_X is the sample standard deviation of observed scores.

The standard error of the mean, on the other hand, is an estimate of the magnitude of sampling error associated with the sample mean in the estimation of the population mean. This expected standard mean of sampling errors of the mean is called the standard error of the mean (SEMn) and is defined as follows:

$$SEMn = \frac{\sigma}{\sqrt{n}} \qquad \text{(Equation 9.5)}$$

where

SEMn = standard error of the mean

σ = standard deviation of the population

n = number of responses in each sample

The more accurate the estimation of the population mean, the smaller the SEMn values will be.

Item Response Theory Conditional SEM

Unlike the classical SEM, the conditional SEM based on Item Response Theory (IRT) is not the same value across test scores. For example, if a person gets either a few or a large number of

items correct (i.e., scores at the extremes of the score distribution), the conditional standard error will be greater in value than it will be if the person gets a moderate number of items correct. This implies that the standard error of measurement depends on the total score (Andrich & Luo, 2004).

Under the Rasch model, the SEM for each person is as follows:

$$\sigma_{\hat{\beta}} = \frac{1}{\sqrt{\sum_{i=1}^{L} p_{vi}(1-p_{vi})}} \qquad \text{(Equation 9.6)}$$

where

v is subscript for a person,

i is subscript for an item,

L is length of the test,

$\hat{\beta}$ is ability estimate, and

p_{vi} is the probability that a person answers an item correctly and is defined as follows:

$$p_{vi} = \frac{e^{\beta_v - \delta_i}}{1 + e^{\beta_v - \delta_i}}$$

where

β_v is person *v*'s ability and δ_i is the item's difficulty.

A confidence band can be used in interpreting the ability estimate. For example, an approximate 68% confidence interval for $\hat{\beta}$ is given by

$$\hat{\beta} \pm SEM$$

Note that the standard error for item difficulty is smallest when the probability of passing is close to the probability of failing. That is, when an item is near the threshold level for many persons in the sample, the standard error is small (Embretson & Reise, 2000).

According to the general consensus in measurement, an aspect for the popularity of IRT methods in analyzing data is based on the fact that classical theory statistics assume equivalency of students and item measurements for all examinees and items in a test. IRT methods allow for the differentiation of varied student and item performances in estimating the reliability of the measurement (Crocker & Algina, 1986). As evidenced by Equation 9.5 above, one reason for the fluctuation in the standard errors of students is that they are a function of the n-counts. As such, the standard errors for each of the ability score estimates are smallest in the middle of the score distribution (where most examinees perform) and greatest for estimates in the extreme where subsequently lower numbers of students perform on a test, and thus produce less precise

estimates. It is for this reason that IRT estimates with individual standard errors at score points, i.e., conditioned on theta (student ability estimates), are preferred to classical SEMs which do not differentiate between the precision of student estimates at different levels of performances.

9.3 Decision Consistency and Accuracy at the Pass (Proficient) Cut Scores

The accuracy of a decision is the extent to which it would agree with the decisions that would be made if each student were tested with all possible parallel forms of the assessments. The consistency of a decision is the extent to which it would agree with the decisions that would be made if the students had taken a different form of the examination, equal in difficulty and covering the same content as the form they actually took. Every test administration will result in some error in classifying examinees. Students can be misclassified in either of two ways. Students who were below the proficiency cut score but were classified (on the basis of the assessment) as being above a cut score are considered to be false positives. Students who were above the proficiency cut score but were classified as being below a cut score are considered to be false negatives. Decision consistency and accuracy are important indications of the quality of an assessment for which performance categories are the primary means of reporting results.[8] Decision consistency and accuracy tables are in part II of this report.

9.4 Inter-Rater Reliability

For the writing assessments, which have an open-ended item that students provide responses on, an additional form of reliability is assessed. Inter-rater reliability investigates the extent to which examinees would obtain the same score if the Virginia SOL writing assessment were scored by different scorers. Inter-rater reliability is calculated as the percent agreement between raters. The metrics tracked and reported are "perfect agreement" and "adjacent agreement." Perfect agreement is when the two independent scorers assign the same score to the same piece of student work. Adjacent agreement is when the two independent scorers assign adjacent score points to the same piece of student work.

10. TEST VALIDITY

As noted in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), "[v]alidity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests." Messick (1989) defined validity as follows:

> Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (p.5)

[8] Part II contains the results of analyses performed to estimate the accuracy and consistency of the decisions for passing (proficient) on the Virginia SOL assessments.

This definition implies that test validation is the process of accumulating evidence to support intended use of test scores. Consequently, test validation is a series of ongoing and independent processes that are essential investigations of the appropriate use or interpretation of test scores from a particular measurement procedure (Suen, 1990).

In addition, test validation embraces all of the experimental, statistical, and philosophical means by which hypotheses and scientific theories can be evaluated. This is the reason that validity is now recognized as a unitary concept (Messick, 1989). Typically, one or more types of validity evidence are desired: face validity, content validity, or construct validity. This report relies mostly on evidence of content and construct validity while touching on the intrinsic rational validity of the Standards of Learning (SOL) program.

10.1 Face Validity

Given that the SOL program is used to measure student achievement on the Virginia SOL, "the validation rests, in part, on the appropriateness of test content, [and] the procedures followed in specifying and generating test content" (Standard 1.6). The SOL tests exhibit evidence of face validity due to the rigor with which the SOL Test Blueprint specifications match the emphases in the SOL Curriculum Frameworks[9] and the involvement of Virginia educators in insuring that each test form matches the blueprint specifications and that each item on each form adequately, appropriately, and fairly addresses the standard of learning being measured. This is somewhat related to the notion of Intrinsic Rational Validity.

10.2 Intrinsic Rational Validity

The process implemented by the Virginia Department of Education (VDOE) to develop and design the SOL program is evidence of the intrinsic rational validity of the SOL assessment. As defined by Ebel (1983), intrinsic rational validity is

> "…evidence that exists as an artifact of the test development process. The evidence is intrinsic, because it is built into the test. It is rational because it is derived from rational inferences about the kind of tasks that will best meet the measurement goal of the assessment" (as cited by Maryland Department of Education, 2004).

10.3 Content Validity

Content validity answers the question, "Does this measure include all the relevant content it is supposed to while excluding irrelevant content?" Content validity is frequently defined in terms of the sampling adequacy of test items. That is, content validity is the extent to which the items in a test adequately represent the domain of items or the construct of interest (Suen, 1990). In educational testing, the state curriculum defines the content that is to be taught and assessed.

[9] The SOL Curriculum Frameworks may be accessed at: http://www.doe.virginia.gov/VDOE/CurriculumFramework .

Consequently, content validity provides judgmental evidence in support of the domain relevance and representativeness of the content in the test (Messick, 1989).

Relation to Content Standards

Each Virginia SOL assessment is built to a specified blueprint which is designed to ensure that each SOL assessment addresses the Commonwealth's content standards within each subject. This blueprint provides guidance on test construction as to the number of items to be used from each content strand within each reporting category. This blueprint is used in each administration so that there is consistency from year-to-year in what is being assessed in relation to the content standards. The blueprint ensures that coverage of the SOL occurs in each assessment.

The content in the SOL Test Blueprint derives directly from the SOL Curriculum Framework. The SOL Curriculum Framework amplifies the Standards of Learning and defines the content knowledge, skills, and understandings that are measured by the Standards of Learning tests. The Curriculum Framework provides additional guidance to school divisions and their teachers as they develop an instructional program appropriate for their students. It assists teachers as they plan their lessons by identifying essential understandings, defining essential content knowledge, and describing the intellectual skills students need to use. This supplemental framework delineates in greater specificity the minimum content that all teachers should teach and all students should learn. This direct relationship between the SOL Curriculum Frameworks, the SOL Test Blueprint, and the SOL assessments lends support to the content validity of the SOL assessments.

Educator Input on Item Development

Test development for Virginia SOL is ongoing and continuous. ETS, content specialists, Virginia educators, Pearson, and VDOE are greatly involved in developing and reviewing test items. The content, data, and bias review committees evaluate and approve all of the items following standardized procedures at multiple stages in the item and test development process prior to the items being placed on any operational test form.

Once embedded field-test items are scored, VDOE and Pearson conduct additional item analysis, and data review meetings are held so that Virginia educators may evaluate the item's performance. Committee members may recommend that the item be eliminated or revised and field-tested again.

Additionally, all items developed and banked for use on operational test forms have gone through multiple rounds of committee review to ensure that they are indeed measuring what they are intended to measure fairly for all students accessing the SOL assessments. Each test form has gone through committee review to ensure that the form is an accurate reflection of the Test Blueprints developed in conjunction with the Commonwealth's assessment and curriculum specialists and that the form is not biased in any way.[10]

[10] See Section 3 for a more detailed explanation of the educator input on item development.

10.4 Construct Validity

Criterion or construct validity answers the question, "Does this measure of this construct behave in ways that are consistent with expectations, underlying theory, or in a similar fashion as *other* measures of this construct?" (Peters, Crossen, and Anderson, 2000). In order to answer this question, VDOE has engaged in ongoing research.

In the content areas and grade levels where there were reasonable matches of content, school pass rates on the SOL tests were previously statistically correlated with national percentile ranks on the *Stanford 9* and/or pass rates on the *LPT.* See Table 10.1 for a summary of the school-level rank order correlations of the SOL and the *Stanford 9* or *LPT*.[11]

Table 10.1 School Level Rank Order Correlations

Comparison	Grade 3	Grade 5	Grade 8	H.S./Grade 11
SOL English: Reading/Literature and Research & *Stanford 9* Reading	.76-.78	.76-.78	.80-.81	.57-.62
SOL Mathematics & *Stanford 9* Math	.72	.76	.82	.71
SOL Algebra I, II, and Geometry & *Stanford 9* Total Mathematics				.53-.71
SOL English: Reading/Literature and Research & Grade 6 *LPT* Reading		.64	.75	
SOL English: Writing & Grade 6 *LPT* Writing		.68	.61	
SOL Mathematics & Grade 6 *LPT* Mathematics			.54	.56

Additionally, to evaluate each of the current assessments for construct validity, the following section shows the results from factor analyses using the spring 2007 assessment data. When linking forms of the assessments across years for the purpose of reporting scores, it is assumed that each Virginia SOL assessment (e.g., Grade 3 mathematics) is, for the most part, measuring a dominant trait or main factor. For example, in the mathematics area, the mathematics assessments should be measuring mathematics competence and not the combined effects of mathematics competence, reading and language proficiency, and the ability to work quickly. Often called the requirement for "test unidimensionality," evidence for the validity of the unifactor or unidimensionality assumption for an assessment can come from performing a factor analysis. Reviewing the "eigenvalue plot" that is obtained from analyzing the correlation matrix formed from the correlations of all pairs of items in each assessment allows one to draw conclusions about the unidimensionality of the assessment. "[Factor analysis] tells us, in effect, what tests or measures belong together – which ones measure virtually the same thing, in other words, and how much they do so." (Kerlinger, 1973, p.659)

[11] Full documentation of the construct validity of the SOL assessments can be found in *Standards of Learning (SOL) tests validity reliability information: Spring 1998 administration* (Virginia Department of Education, 1999). The full tables of correlation statistics may be found in Appendix A of the VDOE (1999) publication.

The SAS procedure PROC FACTOR was conducted on the item response matrix. Only factors with eigenvalues greater than 1 were retained, a criterion proposed by Kaiser (1960). Scree plots were also developed to graphically display the relationship between factors with eigenvalues exceeding 1. Standard practice suggests that when the scree plot appears to level off, it is an indication that the number of significant factors has been reached.

The VASOL assessments are offered in two modes of administration: traditional paper and pencil or online computer-based testing. For each assessment, Tables 10.2-10.7 show the first three eigenvalues for each factor analysis and the value of Divgi's index (Hattie, 1985). This index is the ratio of the difference between the first and second eigenvalues to the difference between the second and third eigenvalues. A value that is greater than 3.0 implies that the test in question is characterized by a dominant first dimension.

The results show that all values of Divgi's index greatly exceed 3.0. This suggests that all core forms across both testing modes and grades are characterized by a dominant primary dimension, and that it is reasonable to apply unidimensional IRT models in the psychometric work that supports them.

Table 10.2 Factor Analyses for Grade 3-8 Reading

Subject	Grade	Core	Mode	N-Count	Eigenvalues			Divgi's Index
					λ_1	λ_2	λ_3	
Reading	3	1	Online	0	NA	NA	NA	NA
			Paper	42388	5.91	0.37	0.30	79.14
		2	Online	3042	6.26	0.56	0.39	33.53
			Paper	19304	6.63	0.53	0.39	43.57
	4	1	Online	0	NA	NA	NA	NA
			Paper	40482	6.49	0.28	0.24	155.25
		2	Online	4373	5.68	0.37	0.31	88.50
			Paper	18765	6.31	0.39	0.32	84.57
	5	1	Online	0	NA	NA	NA	NA
			Paper	39200	6.77	0.56	0.36	31.05
		2	Online	6784	7.46	0.67	0.56	61.73
			Paper	18687	8.40	0.66	0.43	33.65
	6	1	Online	11521	7.39	0.58	0.33	27.24
			Paper	28663	8.02	0.62	0.40	33.64
		2	Online	7396	7.92	0.49	0.33	46.44
			Paper	16283	8.71	0.48	0.32	51.44
	7	1	Online	11750	7.82	0.56	0.45	66.00
			Paper	26555	8.74	0.65	0.47	44.94
		2	Online	10305	7.59	0.66	0.34	21.66
			Paper	16427	8.70	0.70	0.36	23.53
	8	1	Online	14772	8.86	0.64	0.31	24.91
			Paper	19260	8.80	0.65	0.32	24.70
		2	Online	9722	8.34	0.61	0.36	30.92
			Paper	15610	10.10	0.65	0.35	31.50
		3	Online	4385	9.06	0.78	0.37	20.20
			Paper	3540	10.70	0.89	0.44	21.80

Table 10.3 Factor Analyses for Grade 3-8 Mathematics

Subject	Grade	Core	Mode	N-Count	Eigenvalues			Divgi's Index
					λ_1	λ_2	λ_3	
Math	3	1	Online	0	NA	NA	NA	NA
			Paper	44424	6.58	0.57	0.44	46.23
		2	Online	2774	6.94	0.64	0.52	52.50
			Paper	17010	7.06	0.76	0.41	18.00
	4	1	Online	0	NA	NA	NA	NA
			Paper	37598	7.56	.909	.499	16.22
		2	Online	3543	6.84	0.57	0.46	57.00
			Paper	16724	7.35	0.52	0.47	136.60
	5	1	Online	0	NA	NA	NA	NA
			Paper	35394	6.82	0.82	0.65	35.29
		2	Online	5690	6.54	0.91	0.53	14.82
			Paper	14958	6.32	1.04	0.54	10.56
	6	1	Online	10739	7.53	0.79	0.46	20.42
			Paper	24958	8.17	0.79	0.52	27.33
		2	Online	5225	8.78	0.70	0.62	101.00
			Paper	12534	9.08	0.79	0.58	39.48
	7	1	Online	17573	8.74	0.81	0.53	28.32
			Paper	21128	9.36	0.84	0.58	32.77
		2	Online	6539	7.95	0.70	0.40	24.17
			Paper	5934	10.05	0.80	0.46	27.21
	8	1	Online	23194	8.25	0.81	0.57	31.00
			Paper	19052	8.39	0.90	0.62	26.75
		2	Online	8991	7.40	0.74	0.56	37.00
			Paper	4833	9.82	0.95	0.58	23.97
		3	Online	4192	9.42	0.90	0.81	94.67
			Paper	2929	9.76	0.84	0.80	223.00

Table 10.4 Factor Analysis for Grade 3-8 Plain English Mathematics

Subject	Grade	Core	Mode	N-Count	Eigenvalues			Divgi's Index
					λ_1	λ_2	λ_3	
Plain English Math	3	2	Online	0	NA	NA	NA	NA
			Paper	5587	8.50	0.74	0.58	48.50
	4	2	Online	0	NA	NA	NA	NA
			Paper	5870	7.39	0.98	0.49	13.08
	5	2	Online	0	NA	NA	NA	NA
			Paper	6178	7.14	1.31	0.70	9.56
	6	2	Online	1400	7.40	0.72	0.69	222.67
			Paper	3720	7.59	0.67	0.57	69.20
	7	2	Online	2139	6.53	0.67	0.41	22.54
			Paper	2055	5.47	0.73	0.39	13.94
	8	3	Online	2289	7.78	0.90	0.68	31.27
			Paper	1788	8.11	0.83	0.59	30.33

Table 10.5 Factor Analysis for Grades 3, 5, and 8 History and Science

Subject	Grade	Core	Mode	N-Count	Eigenvalues			Divgi's Index
					λ_1	λ_2	λ_3	
History	3	1	Online	2320	5.60	0.47	0.42	102.60
			Paper	41387	5.34	0.43	0.28	32.73
		2	Online	2178	5.80	0.54	0.47	75.14
			Paper	18279	5.91	0.42	0.33	61.00
	8	1	Online	0	NA	NA	NA	NA
			Paper	5568	8.43	0.63	0.36	28.89
Science	3	1	Online	2290	5.14	0.40	0.32	59.25
			Paper	41657	4.95	0.43	0.19	18.83
		2	Online	1756	4.89	0.47	0.33	31.57
			Paper	0	NA	NA	NA	NA
	5	1	Online	5094	4.55	0.28	0.23	85.40
			Paper	38509	4.48	0.25	0.18	60.43
		2	Online	3771	4.44	0.35	0.20	27.27
			Paper	17866	5.04	0.29	0.18	43.18
	8	1	Online	7849	6.96	0.40	0.35	131.20
			Paper	10313	8.51	0.47	0.37	80.40
		2	Online	13057	8.43	0.38	0.26	67.08
			Paper	18201	9.45	0.37	0.23	64.86

Table 10.6 Factor Analysis for Content-Specific History Tests

Subject	Core	Mode	N-Count	Eigenvalues			Divgi's Index
				λ_1	λ_2	λ_3	
Virginia Studies	1	Online	4026	6.06	0.52	0.31	26.38
		Paper	38578	5.96	0.37	0.25	46.58
	2	Online	3486	6.72	0.51	0.33	34.50
		Paper	18124	7.26	0.40	0.30	68.60
U.S. History to 1877	1	Online	10347	7.26	0.52	0.34	37.44
		Paper	16371	7.23	0.55	0.35	33.40
	2	Online	8074	7.45	0.46	0.31	46.60
		Paper	15777	7.75	0.55	0.33	32.73
U.S. History: from 1877 to Present	1	Online	24461	6.74	0.40	0.31	70.44
		Paper	13549	6.81	0.42	0.30	53.25
	2	Online	9497	6.43	0.57	0.33	24.42
		Paper	2749	7.68	0.83	0.51	21.41
Civics and Economics	1	Online	12025	6.98	0.51	0.29	29.41
		Paper	11948	7.40	0.60	0.24	18.89
	2	Online	10737	6.25	0.47	0.38	64.22
		Paper	13657	7.16	0.53	0.26	24.56

Table 10.7 Factor Analysis for High School End-of-Course Tests

Subject	Core	Mode	N-Count	Eigenvalues			Divgi's Index
				λ_1	λ_2	λ_3	
Earth Science	1	Online	21139	7.88	0.47	0.32	49.40
		Paper	0	NA	NA	NA	NA
	2	Online	21392	7.74	0.38	0.30	92.00
		Paper	2084	10.66	0.51	0.39	84.58
	3	Online	11324	7.93	0.43	0.21	34.09
		Paper	0	NA	NA	NA	NA
Biology	1	Online	32094	7.73	0.45	0.36	80.89
		Paper	0	NA	NA	NA	NA
	2	Online	21856	7.76	0.43	0.23	36.65
		Paper	1729	11.17	0.57	0.35	48.18
	3	Online	11162	7.15	0.42	0.31	61.18
		Paper	0	NA	NA	NA	NA
Chemistry	1	Online	23191	7.31	0.70	0.47	28.74
		Paper	0	NA	NA	NA	NA
	2	Online	12632	5.64	0.74	0.71	163.33
		Paper	870	7.63	1.02	0.99	220.33
	3	Online	6348	5.66	0.71	0.63	61.88
		Paper	0	NA	NA	NA	NA

Subject	Core	Mode	N-Count	Eigenvalues			Divgi's Index
				λ_1	λ_2	λ_3	
Algebra I	1	Online	33456	7.72	1.05	0.69	18.53
		Paper	0	NA	NA	NA	NA
	2	Online	19486	7.98	0.97	0.66	22.61
		Paper	2307	10.26	0.97	0.62	26.54
	3	Online	11570	8.20	0.87	0.55	22.91
		Paper	2799	10.58	1.00	0.63	25.89
Geometry	1	Online	28736	7.59	0.54	0.44	70.50
		Paper	0	NA	NA	NA	NA
	2	Online	18716	6.99	0.49	0.30	34.21
		Paper	1904	8.17	0.55	0.42	58.62
	3	Online	9382	7.46	0.66	0.54	56.67
		Paper	2535	6.84	0.81	0.64	35.47
Virginia & United States History	1	Online	28640	9.15	0.66	0.31	24.26
		Paper	0	NA	NA	NA	NA
	2	Online	20294	9.71	0.54	0.34	45.85
		Paper	1965	13.29	0.77	0.48	43.17
	3	Online	10530	10.12	0.61	0.35	36.58
		Paper	0	NA	NA	NA	NA
World History I	1	Online	27366	9.09	0.61	0.36	33.92
		Paper	0	NA	NA	NA	NA
	2	Online	12708	8.58	0.87	0.41	16.76
		Paper	3176	12.78	1.20	0.41	14.66
	3	Online	10510	9.91	0.61	0.41	46.50
		Paper	0	NA	NA	NA	NA
World History II	1	Online	27106	8.73	0.78	0.58	39.75
		Paper	0	NA	NA	NA	NA
	2	Online	15859	9.07	0.71	0.45	32.15
		Paper	2126	11.43	0.99	0.46	19.70
	3	Online	8186	9.43	0.66	0.42	36.54
		Paper	0	NA	NA	NA	NA
World Geography	1	Online	3842	8.82	0.71	0.41	27.03
		Paper	0	NA	NA	NA	NA
	2	Online	10815	8.57	0.74	0.36	20.61
		Paper	981	11.11	1.06	0.76	33.50
	3	Online	2387	9.35	0.59	0.48	79.64
		Paper	0	NA	NA	NA	NA

Subject	Core	Mode	N-Count	Eigenvalues			Divgi's Index
				λ_1	λ_2	λ_3	
English: Reading	1	Online	28654	6.75	0.51	0.29	28.36
		Paper	0	NA	NA	NA	NA
	2	Online	18859	6.48	0.38	0.34	152.50
		Paper	2488	7.18	0.60	0.47	50.62
	3	Online	9121	6.89	0.64	0.31	18.94
		Paper	0	NA	NA	NA	NA
Algebra II	1	Online	24121	7.65	0.61	0.43	39.11
		Paper	0	NA	NA	NA	NA
	2	Online	13703	6.75	0.53	0.47	103.67
		Paper	1917	6.70	0.63	0.54	67.44
	3	Online	7867	7.49	0.76	0.56	33.65
		Paper	0	NA	NA	NA	NA
Plain English Algebra I	3	Online	493	9.24	1.18	1.02	50.38
		Paper	71	10.17	2.62	2.52	75.50

11. ALTERNATE AND ALTERNATIVE ASSESSMENTS

Public Law 105-17, the Individuals with Disabilities Education Act (IDEA) Act of 1997, and its reauthorization, Public Law 108-446, the Individuals with Disabilities Education Improvement Act of 2004, require that states have "established goals for the performance of children with disabilities in the state that...are consistent, to the maximum extent appropriate, with other goals and standards for all students, including those with disabilities, in state- and district-wide assessments with the provision of appropriate and necessary accommodations." For students who cannot participate in state- and district-wide assessments, the law required that state education agencies develop and implement guidelines for their participation in an alternate form of assessment by July 1, 2000. The intent of the federal legislation is to make certain that educational reform efforts include all students. Critical elements in improving education for students with disabilities are promoting high expectations appropriate with their particular needs and ensuring meaningful and effective access to the general curriculum. When schools have high expectations for students with disabilities, guarantee appropriate access to the general curriculum, and provide necessary supports and accommodations, many students can achieve higher standards than society has historically expected.

Federal statutes and regulations specifically address these issues by requiring the development of state performance goals for children with disabilities that must address certain key indicators for success of educational efforts for these students. State-developed goals and indicators must be "consistent, to the maximum extent appropriate, with other goals and standards for children established by the State" [IDEA 1997, Section 612(16)(A)(ii)]. Special education must be viewed as an extension of general education and not as a separate system.

The purpose of the Virginia Grade Level Alternative (VGLA) Program, the Virginia Alternate Assessment Program (VAAP), and the Virginia Substitute Education Program (VSEP) is to evaluate the performance of students with disabilities who are unable to participate in the Virginia Standards of Learning statewide testing program, even with accommodations.[12] The intent of the IDEA, the No Child Left Behind Act of 2001 (NCLB), and these programs is to bring a quality level of participation and accountability to this population of students.

11.1 Virginia Grade Level Alternative (VGLA)

The VGLA assessment program is available for students in grades 3-8 as an alternative assessment for SOL testing. Students with disabilities that prevent them from accessing the SOL test(s) in a content area, even with accommodations, may participate in the VGLA Program. Students who qualify to participate in the VGLA Program are required to demonstrate individual achievement of grade level content standards as presented in the SOL test blueprints for the academic content area in which they are being assessed. Students compile a collection of work samples—a Collection of Evidence (COE)—to demonstrate performance on all *on-grade level* SOL on which they have received instruction. Scoring teams convened by the local school division score student work samples using a rubric developed by the Virginia Department of Education. Scores are entered in an online system that calculates the student's proficiency level.

Identification of students eligible to participate in this program, participation criteria, learner characteristics, and other instructional programming information may be found in the *Virginia Grade Level Alternative Program (VGLA) Procedural Manual* available from the Virginia Department of Education Division of Student Assessment and School Improvement website at http://www.doe.virginia.gov/VDOE/Assessment/home.shtml#VGLA.

Early Development

In response to the requirements of NCLB, the development of the VGLA began in the fall of 2004. VGLA was to:

1. Be administered at the same grade level as required by the State for general education students.
2. Reflect student choice and decision-making.
3. Allow students to demonstrate strengths rather than weaknesses.

Several principles emerged to guide the development of the assessment program. These principles included:

1. Decisions about participation in the VGLA are made collaboratively by the IEP team/504 committee.
2. Schools are accountable and have high expectations for all students.

[12] Limited English Proficient (LEP) students are also able to participate in the VGLA Program.

Collections of Evidence

As noted, teachers of students participating in the Virginia Grade Level Alternative assessment collect information on the performance of students with disabilities through a Collection of Evidence (COE). A COE is an assessment instrument that allows teachers and students to collect data (evidence) and organize this information into a binder to represent individual student performance and achievement for inclusion in the VGLA.

Collections are organized using the SOL Test Blueprint[13] as a guide. At least one piece of evidence must be submitted for **each** SOL listed in the SOL Test Blueprint. Collections of Evidence are prepared by students and teachers working together to assemble a representative sample of student work, data sheets, photographs, and other examples of student performance to evidence individual achievement in the general education curriculum. That a COE must contain a piece of evidence representing performance on each and every SOL measured by the regular SOL assessment is a critical component of its content and construct validity for holding VGLA students to the same standards as regular SOL students.

Scoring Rubric Development

The scoring rubric for the COE focuses on individual student performance and achievement. By design, the rubric provides scorers with a consistent set of standards by which each COE can be reviewed and assigned a score. The rubric helps the scorer assess student performance in English/language arts (reading and writing), mathematics, science, and history/social sciences. The scores assigned to each piece of VGLA evidence can range from "0" to "4":

4 - There is **ample evidence** that the student has demonstrated the skills and knowledge stated in the Standard(s) of Learning being addressed.

3 - There is **adequate evidence** that the student has demonstrated the skills and knowledge stated in the Standard(s) of Learning being addressed.

2 - There is **some evidence** that the student has demonstrated the skills and knowledge stated in the Standard(s) of Learning being addressed.

1 - There is **little evidence** that the student has demonstrated the skills and knowledge stated in the Standard(s) of Learning being addressed.

0 - If evidence submitted **does not show any understanding** of the skills and knowledge listed in the SOL being defended or if **NO evidence is submitted**.

Standard Setting

During the 2008-2009 administration cycle, there were no standard-setting events held.

[13] SOL Test Blueprints may be accessed at: http://www.doe.virginia.gov/VDOE/Assessment/soltests .

11.2 Auditing Student Collections for VGLA

All VGLA Collections of Evidence (COE) are scored by trained Virginia educators during "Local Scoring Events" conducted throughout the state. Following this scoring by local educators, Pearson conducts an audit of 10% of the submitted COE at the Virginia Beach Performance Scoring Site. The auditors receive instruction with the same training materials as the Virginia educators who determine the first score. If the proficiency score assigned during the audit disagrees with the proficiency score assigned at the Local Scoring Event, the collection receives a third score from a Scoring Supervisor or Scoring Director. If this resolution proficiency score does not match either the original first proficiency score OR the second proficiency score, it receives a fourth score from the Scoring Director. This process continues until each audited collection has received two proficiency scores that agree. Scorers are required to take questions about scoring a particular COE and rubric interpretation to their Scoring Supervisor or Scoring Director in every instance.

Selecting Anchor, Practice, and Qualifying Papers

In order to train scorers at the "Local Scoring Events," Performance Scoring Center (PSC) staff pre-screen evidence from COE for rangefinding purposes each administration. These samples are collected during the previous administration's audit process. A sample is chosen to represent:

- a range of school districts;
- variety of grade levels (elementary, middle, high school); and
- all possible rubric scores (low, medium, high).

All rangefinding participants are provided with individual copies of the evidence to be assessed during rangefinding. There are typically three four-day rangefinding sessions.

At the start of the rangefinding meeting, Virginia educators, in conjunction with the PSC staff, begin by reviewing the scoring rubric and addendum, general VGLA scoring rules, and sample pieces of evidence to ensure there is a common understanding of standards and consistency of scoring from year to year. The rangefinding committee is introduced to their tasks of reviewing and scoring rangefinding evidence that will be used in the training of PSC Scorers and local educators. PSC staff members maintain notes and record scores and teacher comments. Committee comments and discussion are used by Pearson staff to aid in training.

Immediately following the rangefinding meeting, VDOE and Pearson meet to finalize and sign off on consensus scores for each COE that will be used in the training events. The Pearson Scoring Directors later add information on the placement of each piece of student evidence in anchor, practice, and qualifying sets. Scoring Directors write annotations that explain the rationale for each score used in training. Anchor, practice, and qualifying sets are approved by VDOE before local Train the Trainer sessions begin, and the PSC audit.

Selecting and Training Auditors

In the selection of candidates for auditing this assessment, priority is given to individuals with teaching experience. Priority is also given to Scorers who have successfully scored the VGLA or another alternate assessment. Regardless of previous experience or education, however, the Scorers are required to meet the project's qualification standards (acceptable scores on a set of qualification papers) and are subject to continual monitoring (i.e., backreading) for quality and accuracy. The PSC verifies and evaluates the experience and credentials of all potential auditors and the Scoring Supervisors. At a minimum, all auditors have a four-year college degree and complete the formal application process including an interview.

Auditors are assigned to one of five content areas (reading, writing, mathematics, science, or history). Training covers all grade levels of the VGLA for each specific subject. First, the VGLA scoring rubric with addendum and scoring rules are presented in context with student Collections of Evidence (COE). The rubric consists of a five-point scale:

- 4 = Ample evidence
- 3 = Adequate evidence
- 2 = Some evidence
- 1 = Little evidence
- 0 = No evidence

The addendum can be accessed at the VDOE Division of Student Assessment and School Improvement website at http://www.doe.virginia.gov/VDOE/Assessment/home.shtml

Next, a subject-specific anchor set of individual pieces of evidence, consisting of all training issues, is introduced to auditors. Then, two practice sets are used to give the auditors the opportunity to practice scoring. Finally, a qualifying set is administered to auditors to determine if they have fully grasped the scoring criteria and rules. Each auditor is required to attain at least a score of 70% on one of two qualifying sets. If 70% or higher is not attained on the first qualifying set, the trainee scores a second qualifying set to attempt to achieve the 70%.

Scoring Supervisors are trained before the rest of the auditors. When possible, Scoring Supervisors have previous experience scoring alternate assessments. Scoring Supervisors are chosen based on their ability to score accurately and to communicate the rubric standards to Scorers. The Scoring Supervisor is responsible for supervising the auditors and monitoring their performance.

Backreading

Backreading is one of the primary responsibilities of Scoring Supervisors. Scoring Supervisors "read behind" the auditors, reviewing a random sample of the scores assigned by each of them to ensure accuracy. Backreading results are documented and recorded by the Supervisor on backreading tally forms. Backreading continues throughout the scoring of the project. Pearson's immediate backreading process helps identify individual trends and tendencies and is the foundation for the individual feedback and retraining provided.

Each morning, the auditors review their anchor sets and general scoring decisions. At the end of each day of scoring, the Scoring Supervisor meets with the Scoring Director to discuss any scoring decisions or issues that are causing inconsistencies. During the audit scoring, collections are copied to be used for the following year's rangefinding meetings.

11.3 Virginia Alternate Assessment Program (VAAP)

The Virginia Alternate Assessment Program (VAAP) is designed to evaluate the performance of students with significant cognitive disabilities. The VAAP is available to students in grades 3 through 8 and students in grade11 who are working on academic standards that have been reduced in complexity and depth. This content is derived from the Standards of Learning (SOL) and is referred to as the Aligned Standards of Learning (ASOLs).

Appropriate content level standards are chosen based on reduced complexity in skill/knowledge statements from each reporting category in each content area across grade levels. One ASOL is selected from each reporting category expressed from the content area and grade level being assessed.

The VAAP is an evidence-based design that permits students to demonstrate their knowledge and skills through various types of evidence including work samples, anecdotal records, interviews, captioned photographs, videotapes and audiotapes. The collection of student work known as a Collection of Evidence (COE) allows teachers and students to collect data (evidence) and organize this information into a binder to represent student performance and achievement on the ASOLs selected for inclusion in the VAAP. Flexible in design and construction, the COE allows all students being assessed through the VAAP the ability to demonstrate proficiency in their selected ASOL without adhering to a specific type of data or assessment tool.

The student's teacher collects evidence of individual achievement throughout the school year that reflects student performance of the selected ASOL. The collections should be student focused with teachers acting as coordinators and facilitators of student performance, based on appropriate levels of communication and participation skills exhibited by individual students. Collections should reflect student performance on the ASOLs.

Identification of students eligible to participate in the VAAP, participation criteria, learner characteristics, and other instructional programming information may be found in the *Virginia Alternate Assessment Program (VAAP)Implementation Manual,* which is available on the Virginia Department of Education's website at: http://www.doe.virginia.gov/VDOE/Assessment/home.shtml.

Early Development

The VAAP development process began fall 1999. During the initial development, the Virginia Alternate Assessment Steering Committee was organized, consisting of fifteen Virginia educators and education professionals from across the Commonwealth. The steering committee

was instrumental in designing and implementing the VAAP from its inception. The steering committee members were charged with creating an assessment system that would:

1. Be administered at the same grade level as required by the State for general education students.
2. Reflect student choice and decision-making.
3. Allow students to demonstrate strengths rather than weaknesses.
4. Demonstrate skills in multiple settings.
5. Use technology or assistive technology when appropriate.

Serving as a research, development, and advisory group, the steering committee was responsible for articulating program philosophy, linking the assessment to the SOL, identifying what information should be collected to measure student achievement, and drafting the scoring criteria.

As a result of the steering committee's work, several principles emerged to guide the development of the VAAP. These principles include:

1. The VAAP is designed for students who are pursuing a functional curriculum regardless of their educational placement (e.g., general education classroom, special education classroom, hospital, homebound, private school, state-operated program).
2. Decisions about participation in the VAAP are made collaboratively by the IEP team.
3. Students participating in the VAAP must have access to and show progress in the general education curriculum to improve the student's quality of life and prepare students for employment and independent living.
4. Student performance in a variety of settings with social interactions and in natural context will be based on multiple sources of data.
5. The VAAP must yield reliable and valid information that leads directly to student learning and improved instruction.
6. The VAAP will follow nondiscriminatory practices and will be sensitive to issues of cultural competence.
7. Student performance on the Life Skills Strands and Performance Indicators and access to the Delivery Practices are viewed as equally important in improving the student's quality of life and in preparing them for employment and independent living.
8. The VAAP will parallel the state- and division-wide assessment to the greatest extent possible.
9. Schools will be accountable and have high expectations for all students.

Summary of Implementation

Field testing occurred during the 1999-2000 school year. The Virginia Department of Education began full implementation of the VAAP during the 2000-2001 school year. The initial VAAP required a collection of student work in reading, mathematics, science and history/social science

that represented student performance and achievement on IEP goals linked to the Standards of Learning. Collections were scored on a 4 point rubric (0 – 3) on the following dimensions: Student Performance, Linkage to the Standards of Learning, Variety of Settings / Social Interactions, Context of Instructional Delivery, and Level of Supports.

Students and teachers began collecting evidence of student performance in October 2000. This evidence was organized into collections of evidence and submitted for scoring to an external contractor through the 2004-2005 administration.

Changes in federal requirements mandated that all general grade level state standards be accessible to all students, including students with the most significant cognitive disabilities. IDEA 2004 acknowledged the inappropriateness of assessing students with significant cognitive impairments using traditional assessments based on state established content standards. The law recognized that this population of learners had instructional needs beyond those of the general population and allowed the development of alternate assessments based on aligned standards.

In 2005-2006, the VAAP was revised in response to the new federal requirements. Under the new mandates VAAP participants were required to demonstrate individual achievement on state established content standards reduced in complexity, Aligned Standards of Learning, and not simply on IEP goals or objectives. In addition, the revised VAAP also included communication skills related to the ASOLs demonstrated across the domains of context, settings, socialization, and communication support.

Beginning in the 2005-2006 administration, VAAP collections of evidence were scored by local school divisions in local scoring events. Scorers for the local events were trained in the scorer process using materials and resources provided by the Virginia Department of Education. Scores were entered into an online system for the calculation of proficiency levels. Ten percent of all scores entered into the online system were randomly selected for audit by an external contractor.

In 2006-2007, the evaluation of communication skills was eliminated with the revised VAAP focusing on individual achievement of ASOLs in the content areas of reading, mathematics, science, and history/social science.

Aligned Standards of Learning

ASOLs are Virginia's approach to providing performance and achievement tools to students with significant cognitive disabilities and assessing them on the grade level content standards expressed in the general curriculum SOL for each reporting category and grade level. The aligned standards are the essential skills and/or knowledge expressed in the regular SOL tests reduced in complexity, or modified to reflect prerequisite skills.

The ASOLs were developed by groups of Virginia educators who reviewed the SOL in each content area and identified the SOL most appropriate for VAAP students and then organized them into reporting categories that match the SOL reporting categories. These reporting categories were then used to develop the ASOL Test Blueprints for each content and grade level to be assessed (reading, mathematics, history/social science, and science in grades 3-8 and 11).

Each Collection of Evidence (COE) must address a single standard from each reporting category – the selected ASOLs may be from any grade level. This means that a COE assessing reading will have two (2) ASOLs, mathematics will have five (5), and science and history will both have four (4).

Scoring Rubric Development

The scoring rubric for the VAAP assessment focuses on individual student performance and achievement. By design, the rubric provides scorers with a consistent set of standards by which each Collection of Evidence can be reviewed and assigned a score. The rubric helps the scorer assess student performance in English/language arts, mathematics, science, and history/social sciences. Scores can range from "1" to "4" – there is no earned "0." See Table 11.2 below for the VAAP scoring rubric.

Table 11.2 VAAP Scoring Rubric

Point	Descriptor
1	There is *little evidence* that the student has demonstrated the skills and knowledge stated in the Aligned Standard(s) of Learning being addressed.
2	There is *some evidence* that the student has demonstrated the skills and knowledge stated in the Aligned Standard(s) of Learning being addressed.
3	There is *adequate evidence* that the student has demonstrated the skills and knowledge stated in the Aligned Standard(s) of Learning being addressed.
4	There is *ample evidence* that the student has demonstrated the skills and knowledge stated in the Aligned Standard(s) of Learning being addressed.

11.4 Auditing Student Collections for VAAP

All VAAP Collections of Evidence (COE) are scored by trained Virginia educators during "Local Scoring Events" conducted throughout the state. Following this scoring by local educators, Pearson conducts an audit of 10% of the submitted COE at the Virginia Beach Performance Scoring Site. The auditors receive instruction with the same training materials as the Virginia educators who had performed the first score. If the proficiency score assigned during the audit disagrees with the proficiency score assigned at the Local Scoring Event, the collection receives a third score from a Scoring Supervisor or Scoring Director. If this resolution proficiency score does not match either the original first proficiency score OR the second proficiency score, it receives a fourth score from the Scoring Director. This process continues until each audited collection has received two proficiency scores that agree. Scorers are required to take questions about scoring a particular COE and rubric interpretation to their Scoring Supervisor or Scoring Director in every instance.

Selecting Anchor, Practice, and Qualifying Papers

In order to train scorers at the "Local Scoring Events," Performance Scoring Center (PSC) staff pre-screen evidence from COE for rangefinding purposes each administration. These samples are collected during the previous administration's audit process. A sample is chosen to represent:

- a range of school divisions;
- variety of grade levels (elementary, middle, high school); and
- all possible scores (low, medium, high).

At the start of the rangefinding meeting, Virginia educators, in conjunction with the PSC staff, begin by reviewing the scoring rubric and addendum, general VAAP scoring rules, and sample pieces of evidence to ensure there is a common understanding of standards and consistency of scoring from year to year. The rangefinding committee is introduced to their tasks of reviewing and scoring rangefinding evidence that will be used in the training of PSC Scorers and local educators. PSC staff members maintain notes and record consensus scores, teacher comments, and discussions of COE. Committee comments and discussion are used by Pearson staff to aid in training.

Immediately following the rangefinding meeting, VDOE and Pearson meet to finalize and sign off on consensus scores for each COE that will be used in the training events. The Pearson Scoring Directors later add information on the placement of each piece of student evidence in anchor, practice, and qualifying sets. Scoring Directors write annotations that explain the rationale for each score of each piece of evidence used in training. Anchor, practice, and qualifying sets are approved by VDOE before Train the Trainer sessions begin at the "Local Scoring Events" and the PSC audit.

Selecting and Training Auditors

In the selection of candidates for auditing this assessment, priority is given to individuals with teaching experience. Priority is also given to auditors who have successfully scored the VAAP or another alternate assessment. Regardless of previous experience or education, however, the auditors are required to meet the project's qualification standards (acceptable scores on a set of qualification papers) and are subject to continual monitoring (i.e., backreading) for quality and accuracy. The PSC verifies and evaluates the experience and credentials of all potential auditors and the Scoring Supervisors. At a minimum, all auditors have a four-year college degree and complete the formal application process including an interview.

Auditors were trained to score all grade levels in reading, mathematics, science, and history content areas. First, the VAAP scoring rubric with addendum and scoring rules are presented in context with student Collections of Evidence (COE). The rubric consists of five- point scale:

- 4 = Ample evidence
- 3 = Adequate evidence
- 2 = Some evidence
- 1 = Little evidence
- 0 = No evidence

Following training on the scoring rubric, a subject-specific anchor set of individual pieces of evidence is introduced to auditors. Then, two practice sets are used to give the auditors the opportunity to practice scoring. Next, to simulate actual scoring of a COE, a complete COE, containing all four content areas, is presented for additional practice. Before qualification, the auditors review a set of evidence that presented issues that should be reviewed by the scoring supervisor. Finally, a qualifying set is administered to the scorers to determine if they have fully grasped the scoring criteria and rules. Each trainee is required to attain at least a score of 65% on one of two qualifying sets. If 65% or higher is not attained on the first qualifying set, the trainee scores a second qualifying set to attempt to achieve the 65%.

The Scoring Supervisors are trained with the rest of the auditors. When possible, Scoring Supervisors have previous experience scoring alternate assessments. Scoring Supervisors are chosen based on their ability to score accurately and to communicate the rubric standards to Scorers. Scoring Supervisors are responsible for supervising the auditors and monitoring their performance.

Backreading

Backreading is one of the primary responsibilities of Scoring Supervisors. Scoring Supervisors "read behind" the auditors, reviewing a random sample of the scores assigned by each of them to ensure accuracy. Backreading results are documented and recorded by the Supervisor on backreading tally forms. Backreading continues throughout the scoring of the project. Pearson's immediate backreading process helps identify individual trends and tendencies and is the foundation for the individual feedback and retraining provided.

Each morning, the auditors review their anchor sets and general scoring decisions. At the end of each day of scoring, the Scoring Supervisor meets with the Scoring Director to discuss any scoring decisions or issues that are causing inconsistencies. Each day the Scoring Director and the Assistant Scoring Director review the overturned rate to see if any auditor had a substantially higher incidence of overturned scores. If this occurs the scoring director works with that auditor for retraining and is backread more frequently.

11.5 Virginia Substitute Evaluation Program (VSEP)

The Virginia Substitute Evaluation Program (VSEP) is a means of verifying high school credits for graduation using student work samples. The VSEP is available as an alternative assessment for students who by the nature of their disability are unable to participate in the regular SOL assessments even with testing accommodations. Under the VSEP, students create a collection of work samples—a Course Work Compilation (CWC)—to demonstrate the *on-grade level* content standards they have learned while taking certain courses in high school. The VSEP requires that students demonstrate proficiency on all the standards addressed within an assessment based on the blueprints for that particular test. Collections of student work samples are evaluated by scoring teams convened by the Department. Currently, this program is only available for courses carrying verified credits and/or for the literacy and numeracy assessments for the Modified Standard Diploma.

Identification of students eligible to participate in this program, participation criteria, learner characteristics, and other instructional programming information may be found in the *Virginia Substitute Evaluation Program (VSEP) Procedural Manual* available at the Virginia Department of Education Division of Student Assessment and School Improvement website at http://www.doe.virginia.gov/VDOE/Assessment/home.shtml#Virginia_Substitute_Evaluation_Program.

11.6 Scoring VSEP

Pearson scored VSEP Course Work Compilations (CWC) at the Performance Scoring Site. Each CWC was scored independently by two Scorers. If the first and second Scorer did not agree on the proficiency level for a CWC, the compilation received a third score from a Scoring Supervisor or Scoring Director. If this resolution proficiency score did not match either the original first proficiency score OR the second proficiency score, it received a fourth score from the Scoring Director. This process continued until each collection had received two proficiency scores that agreed. Scorers were required to take any questions about scoring a particular CWC to their Scoring Supervisor or Scoring Director.

Selecting Anchor, Practice, and Qualifying Papers

In order to train Scorers, the Performance Scoring Center (PSC) staff pre-screen evidence from CWC for rangefinding purposes each administration. These samples are collected during the previous administration's audit process. A sample is chosen to represent:

- variety of secondary courses with end-of-course assessments and grade 8 reading and mathematics
- all possible score levels (low, medium, high)

All rangefinding participants are provided with individual copies of the evidence to be assessed during rangefinding.

At the start of the rangefinding meeting, Virginia educators, in conjunction with the PSC staff, begin by reviewing the scoring rubric and addendum, general VGLA scoring rules, and sample pieces of evidence to ensure there is a common understanding of standards and consistency of scoring from year to year. The rangefinding committee is introduced to their tasks of reviewing and scoring rangefinding evidence that will be used in the training of PSC Scorers and local educators. PSC staff members maintain notes and record consensus scores, teacher comments, and discussions of CWC. Committee comments and discussion are used by Pearson staff to aid in training.

Immediately following the rangefinding meeting, VDOE and Pearson meet to finalize and sign off on consensus scores for CWC samples of students' work that will be used in the training events. The Pearson Scoring Directors later add information on the placement of each piece of student evidence in anchor, practice, and qualifying sets. Scoring Directors write annotations that

explain the rationale for each score of each piece of evidence used in training. Anchor, practice, and qualifying sets are approved by VDOE before scoring sessions begin at the PSC.

Selecting and Training Scorers

In the selection of candidates for scoring this assessment, priority is given to individuals with teaching experience. Priority is also given to Scorers who have successfully scored the VSEP or another alternate assessment. Regardless of previous experience or education, however, the Scorers are subject to continual monitoring (i.e., backreading) for quality and accuracy. The PSC verifies and evaluates the experience and credentials of all potential Scorers and Scoring Supervisors. At a minimum, all Scorers have a four-year college degree related to the VSEP subject that they score. All Scorers complete a formal application process including an interview.

Scorers are assigned to one of five content areas (reading, writing, mathematics, science, or history). Training covers all grade levels of the VSEP for each specific subject. First, the VSEP scoring rubric with addendum and scoring rules are presented in context with student CWC. The rubric consists of a five-point scale.

- 4 = Ample evidence
- 3 = Adequate evidence
- 2 = Some evidence
- 1 = Little evidence
- 0 = No evidence

Next, a subject-specific anchor set of individual pieces of evidence, consisting of all training issues, is introduced to Scorers. Then, a practice set is used to give the Scorers the opportunity to practice scoring.

Scoring Supervisors are trained before the rest of the Scorers. When possible, Scoring Supervisors have previous experience scoring alternate assessments. Supervisors are chosen based on their ability to score accurately and to communicate the rubric standards to Scorers. The Scoring Supervisor is responsible for supervising the Scorers and their performance.

Backreading

Backreading is one of the primary responsibilities of Scoring Supervisors. Scoring Supervisors "read behind" the Scorers, reviewing a random sample of the scores assigned by each of them to ensure accuracy. Backreading results are documented and recorded by the Supervisor on backreading tally forms. Backreading continues throughout the scoring of the project. Pearson's immediate backreading process helps identify individual trends and tendencies and is the foundation for the individual feedback and retraining provided.

Each day before scoring, the Scorers review their anchor sets and the VSEP general scoring decisions. At the end of each day of scoring, the Scoring Supervisor meets with the Scoring Director to discuss any issues that are causing inconsistencies. VDOE is consulted about difficult scoring decisions.

Appeals Process

The primary purpose of the appeals process is to provide an additional step to ensure that the score assigned to the student's VSEP Course Work Compilation is an accurate representation of the student's achievement. A school division may request that a student's CWC be rescored if:

- the student failed the test, **AND**
- there is evidence that the student's VSEP Course Work Compilation should have received a higher score. Evidence of this requires that at least two people familiar with the rubric used to score the CWC review the CWC and agree that it should have received a higher score.

Appeals to rescore a student's CWC may be initiated by parents or by school personnel. All requests for appeals must be reviewed and approved by the school division before being submitted. Appeals to rescore such CWC should be approved by the school division only if the reviewers agree that the CWC should have received a higher score according to the rubric.

11.7 Alternate and Alternative Assessments: Scores and Reports

The Reports provided for the Alternate and Alternative Assessments are similar to those provided for the SOL Assessments. The Reports provided are shown in Table 11.3.

Table 11.3 Alternate and Alternative Assessment Reports Provided

Student Reports	Summary Reports
Student Data Extract by Division	Summary Record Extract by School
Student Data Extract by School	Summary Record Extract by Division
Report to Parents by School	Summary Report by School
Report to Parents by Group	Summary Report by Division
Report to Parents by School by Division	Summary Report by School by Division
Student Performance Report by School	
Student Performance Report by School by Division	
Student Performance Report Self-Adhesive Label by School	

11.8 Reliability of the Alternate and Alternative Assessments

For the alternate and alternative assessments, which have only performance-based items, reliability is assessed in terms of inter-rater reliability. Inter-rater reliability investigates the extent to which examinees would obtain the same performance level if the COE or CWC had been scored by different scorers. Inter-rater reliability is calculated as the percent agreement between raters. The metrics tracked and reported are "perfect agreement" and "adjacent agreement." Perfect agreement is when the two independent scorers assign the same

performance level to the same collection of student work. Adjacent agreement is when the two independent scorers assign adjacent proficiency levels to the same collection of work.

VGLA Inter-Rater Reliability

Pearson conducted audits of 4482 VGLA Collections of Evidence (approximately 10% of VGLA collections submitted). Approximately 69% of the scores assigned in the Virginia school divisions were upheld in audit where the performance levels assigned by the local scorer and the PSC auditor had exact agreement on the first read.

VAAP Inter-Rater Reliability

Pearson conducted audits of 778 VAAP collections of evidence containing 2692 subjects (approximately 10% of VAAP collections submitted). Approximately 86% of the scores assigned in the Virginia school divisions were upheld in audit where the performance levels assigned by the local scorer and the PSC auditor had exact agreement on the first read.

VSEP Inter-Rater Reliability

Each VSEP Course Work Compilation was read and scored by two independent readers, who evaluated each piece of evidence on a 1 to 4 point scale or rubric. The performance level was derived from the scores readers assigned. When the two performance levels were the same, the performance levels were in *exact agreement*. If the performance levels were not in agreement, a scoring supervisor scored the CWC. If the scoring supervisor's performance level matched either of the previous scorers the higher of the scores associated with that performance level was reported. Pearson scored all 173 VSEP collections with 100% of the collections receiving second scores. First and second reader agreement was 87%. First, second and third reader agreement was 99%. All collections were scored until two performance levels were the same.

11.9 VGLA Validity

The VGLA program demonstrates evidence of *intrinsic rational validity* in the way it was developed and implemented by a wide variety of specialists who kept focused on the curriculum frameworks and the population being served. The program demonstrates *content validity* in that it measures each and every standard of learning listed on the SOL Test Blueprints which are themselves a reflection of the Curriculum Frameworks developed by content specialists and Virginia educators. The program exhibits *construct validity* by showing that the SOL (and indirectly the VGLA) assessments are correlated with other nationally recognized assessments and also by the fact that the performance standards that are aligned with the SOL assessments result in a similar pattern of impact.

Intrinsic Rational Validity

The process implemented by the Virginia Department of Education to develop and design the VGLA program, described in Section 11.1 of this report, is evidence of the intrinsic rational validity of the VGLA assessment. VDOE was conscientious in involving content specialists, alternate assessment specialists, and measurement experts to ensure that the program was developed and implemented appropriately given the population of students being assessed and the federal requirements that the program must meet. Virginia educators, local directors, and other state's directors were involved in the process throughout and provided feedback and guidance on how the VGLA program was to be designed and implemented. Such stakeholder involvement helped to ensure that the results of the VGLA assessments would be viewed as meaningful and important to teachers.

Content Validity

The content in the SOL Test Blueprint derives directly from the SOL Curriculum Framework. The SOL Curriculum Framework amplifies the Standards of Learning and defines the content knowledge, skills, and understandings that are measured by the Standards of Learning tests. The Curriculum Framework provides additional guidance to school divisions and their teachers as they develop an instructional program appropriate for their students. It assists teachers as they plan their lessons by identifying essential understandings, defining essential content knowledge, and describing the intellectual skills students need to use. This supplemental framework delineates in greater specificity the minimum content that all teachers should teach and all students should learn.

The design of the Standards of Learning assessment program requires that all Virginia school divisions prepare students to demonstrate achievement of the standards for elementary and middle school by the grade levels tested; therefore, the intention is that all VGLA students are instructed in all areas covered by the framework. This counters a unique threat to the validity of alternate assessments mentioned by Schafer (2005) – the adequacy of the alternate assessment learning domain and the availability of a pathway to access the regular assessment learning domain. In the case of the VGLA assessment, the learning domains (i.e., the curriculum frameworks and assessment blueprints) are identical.

This direct relationship between the SOL Curriculum Frameworks, the SOL Test Blueprint, and the VGLA COE lends support to the content validity of the VGLA assessments. Additionally, the scoring rubric, range of types of evidence that may be submitted, and available accommodations help to ensure that the VGLA is fair to all students of all abilities and limitations which is essential to claims of validity.

Construct Validity

While the construct/criterion validity of the regular SOL assessments is established directly by the evidence of the previously presented rank order correlations, the construct/criterion validity

of the VGLA must be established indirectly through its link to the regular SOL assessment. This indirect link was previously established and reported in the 2006-2007 Technical Report.

11.10 VAAP Validity

The VAAP program demonstrates evidence of *intrinsic rational validity* in the way it was developed and implemented by a wide variety of specialists who kept focused on the Aligned Standards of Learning and the population being served. The program demonstrates *content validity* in that it flexibly allows the performance of VAAP students to be measured on the ASOLs listed on the Aligned SOL Test Blueprints which are themselves a reflection of the SOL and SOL Curriculum Frameworks developed by content specialists and Virginia educators.

Intrinsic Rational Validity

The process implemented by the Virginia Department of Education to develop and design the VAAP program, described in Section 11.2 of this report, is evidence of the intrinsic rational validity of the VAAP assessment. The VDOE was conscientious in involving content specialists, alternate assessment specialists, and measurement experts to ensure that the program was developed and implemented appropriately given the population of students being assessed and the federal requirements that the program must meet. Virginia educators, local directors, and other state directors were involved in the process throughout and provided feedback and guidance on how the VAAP program was to be designed and implemented. Such stakeholder involvement helped to ensure that the results of the VAAP assessments would be viewed as meaningful and important to teachers and parents.

Content Validity

The VAAP Collections of Evidence (COE) are directly based on the ASOLs which were derived from the SOL, and evidence of ASOL from each reporting category listed on the blueprint is intended to be included in each student's COE. This direct relationship between the ASOLs, the SOL Curriculum Frameworks, the SOL Test Blueprint, and the VAAP COE lends support to the content validity of the VAAP assessments. Additionally, the scoring rubric, range of types of evidence that may be submitted, and available accommodations help to ensure that the VAAP is fair to all students of all abilities and limitations which is essential to claims of validity.

12. RESOURCES

In addition to the information presented in this technical manual (Part I and Part II), other resources are available that provide specific details on a variety of topics pertaining to the Virginia SOL assessments. These include administration manuals, released versions of the paper SOL assessments, and Electronic Practice Assessment Tools (ePAT) applications that may be accessed for further documentation and information.

12.1 Administration Manuals

Table 12.1 lists the *Examiner Manuals* and *Test Implementation Manuals* (TIMs) for the 2008-09 testing cycle that were printed, distributed, and posted online.

Table 12.1 VASOL 2009-2009Administration Manuals

Administration	Manual
Fall 2008	2008-09 Grades 5, 8 & EOC Writing Examiner's Manual
	2008-09 EOC Reading Examiner's Manual
	2008-09 EOC History Examiner's Manual
	2008-09 EOC Science Examiner's Manual
	2008-09 EOC Mathematics Examiner's Manual
	2008-09 Grades 6,7 & 8 Non-Writing Examiner's Manual
	2008-09 Content-Specific History Examiner's Manual
	Fall 08 Writing Test Implementation Manual (TIM)
	Fall 08 Non-Writing TIM
Spring 2009	Spring 09 Grade 3 Non-Writing Examiner's Manual
	Spring 09 Grade 4 & 5 Non-Writing Examiner's Manual
	Spring 09 Writing TIM
	Spring 09 Non-Writing TIM
Summer 2009	Summer 09 Writing TIM
	Summer 09 Non-Writing TIM
2008-2009 Administrations	VGLA Manual
	VSEP Manual
	VAAP Manual
	V-Programs Administrator's Manual

12.2 Released Tests

Each spring, the Virginia Department of Education releases a sample set of Standards of Learning (SOL) tests that were administered to Virginia public school students during the previous spring test administration. The sets of released tests are not inclusive of all SOL tests administered during the previous year; however, the tests are representative of the content and skills assessed by the SOL assessment program.

The following 2008 SOL tests were released in March 2009: Grade 3 Reading, Grade 4 Reading, Grade 5 Reading, Grade 6 Reading, Grade 7 Reading, Grade 8 Reading, End-of-Course English: Reading (2002), Grade 3 Mathematics, Grade 4 Mathematics, Grade 5 Mathematics, Grade 6 Mathematics, Grade 7 Mathematics, Grade 8 Mathematics, Algebra I, Geometry, Algebra II (2001 Revised), Grade 3 Science, Grade 5 Science, Biology, Chemistry, Grade 5 Writing, and Grade 8 Writing.

The following 2008 SOL tests will not be released in Spring 2009 in order to protect the security of the SOL item bank: Grade 8 Science, Earth Science, End-of-Course Writing, History and Social Science (all levels).

12.3 Electronic Practice Assessment Tools

End-of-Course (EOC) and middle school training tests are available on the PEMSolutions Training Center for students who will be participating in online testing. The training test will provide students an opportunity to practice navigating through the test and become familiar with the tools and buttons within the test delivery application, TestNav™.

The Electronic Practice Assessment Tools (ePAT) application is a stand-alone program without using an Internet connection after initial download. ePAT prepares students for online testing by simulating an SOL Web-based assessment via the TestNav™ application using released SOL test items. The student will experience the ePAT application in much the same way as an SOL Web-based assessment.[14]

[14] The ePAT applications may be accessed by visiting the Virginia home page (http://pearsonaccess.com/va) or the ePAT home page (http://www.pearsonaccess.com/va/training.center.htm).

PART II: STATISTICAL SUMMARIES FOR 2008-09

1. OVERVIEW OF STATISTICAL SUMMARIES

This section contains an overview of the statistical summaries for the spring 2009 administration of the Virginia SOL assessment.

Analyses are provided for the writing assessments in grades 5, 8 and high school end-of-course, and the multiple-choice assessments for grades 3 through 8 and high school end-of-course. For the writing assessments, analyses are provided for each combination of multiple-choice core and writing prompt for the Main and Alternate administration. Analyses for the multiple-choice assessments for both the Core 1 (Main) and Core 2/3 (Alternate) forms of the assessments are included.

1.1 Administration Results

Three sets of tables are included in the Administration Results section. The first set shows the percentage of students that participated in online or paper administration in the spring 2009 administration. The second set shows the percentages of students in the Proficient and Advanced Proficient performance levels and the overall pass rate for each of the SOL assessments in the spring 2009 administration. The last set shows the raw score summary statistics for all newly constructed SOL assessments across both modes of administration (online and paper-pencil) in the spring 2009 administration (Section 2.1).

1.2 Reliability Estimates for Multiple-Choice Assessments

In order to ensure that quality tests are in place, analyses of the reliability of these new tests, including Plain English Mathematics, were performed. This section focuses on the analyses that were done for all newly constructed cores (1, 2, and 3) of the SOL multiple-choice assessments across both modes of administration (online and paper-pencil) and by subgroups based on gender and ethnicity (white and black) for non-writing tests.

Alpha reliability coefficients can range from 0 to 1 with values greater than .70 considered acceptable (Nunnaly, 1978). Students not reporting their gender or ethnicity were excluded from the subgroup analyses. Results of the reliability analyses for the spring 2009 administration are presented in Section 2.2. These tables show the reliabilities for each grade and subject across core forms and include:

- Numbers of students
- Mode of administration
- *Cronbach's Alpha* internal consistency reliability estimate

1.3 Reliability Estimates for Writing Assessments

The reliability of all the writing assessments was estimated using stratified alpha. Tables in Section 2.3 present stratified alpha for each combination of multiple-choice core and writing prompt for the Main and Alternate administration in grades 5, 8, and high school end-of-course. Analyses were done for the overall group and by subgroups based on gender and ethnicity (white and black).

Each Writing test prompt was read and scored by two independent readers on a 1 to 4 point scale or rubric. When the two readers assigned the same score to a student's paper, the scores were in *exact agreement*. Scores that differed by exactly one score point were *adjacent*. Scores that differed by two or more score points were *non-adjacent*. The room directors reviewed all of the non-adjacent papers before a final score was assigned. These tables cover daily and cumulative results, for each new prompt administered in spring 2009, and include:

- Numbers of students for which the writing domain inter-rater reliabilities were calculated
- Percentages of papers that were in exact agreement, adjacent, or non-adjacent

1.4 Decision Consistency and Accuracy Indices

Every test administration will result in some error in classifying examinees. Accuracy and consistency of the decisions are important indications of the quality of the assessment for which performance categories are the primary means of reporting results. Section 2.4 presents the results of decision consistency and accuracy analyses for the SOL assessments newly constructed for the spring 2009 administration. The analyses apply the methods outlined and implemented in Livingston and Lewis (1995), Haertel (1996), and Young and Yoon (1998). The analyses were done for paper and online administration separately.

For each SOL multiple-choice and writing assessment, these tables include the numbers of students and the proportion of:

- Accurate classifications
- False positives
- False negatives
- Consistent classifications

Although there is no general rule to determine the acceptable levels of decision accuracy and consistency needed for educational assessments, the Virginia SOL assessments have decision accuracy and consistency levels comparable to those that are reported in the Livingston and Lewis paper that describes the procedure. As expected, decision accuracy is generally higher than decision consistency.

1.5 Raw Score to Scale Score Conversion Tables and Conditional SEM

Section 2.5 contains tables for the raw score to scale score (SS) conversions and the conditional standard error of measurement (SEM) at each scale score level for all newly constructed cores (1, 2, and 3) of the multiple-choice assessments across both modes of administration (online and

paper-pencil). For the writing assessments, these tables are provided for each combination of multiple-choice core and writing prompt for the Main and Alternate administration.

2. SPRING 2009 STATISTICAL SUMMARY

2.1 Administration Results

2.1.1 Participation by Mode of Administration

The following tables show the number of tests administered in the online and paper modes of administration. Each table shows the grade and subject area of the test, total number of valid tests administered, and the percent of tests that were administered online and on paper (if a given core is **not** administered in a particular mode, then these cells are shaded).

Table 2.1.1.1 Percent of Tests Taken by Mode: Grades 3-8

Grade	Subject	Total Number	Mode	
			Online (%)	Paper (%)
3	Reading	87,035	16	84
	Math	83,376	16	84
	Plain English Math	6,632		100
	Science	88,480	20	80
	History	88,299	22	78
4	Reading	85,645	20	80
	Math	81,792	19	81
	Plain English Math	5,989		100
5	Reading	85,831	22	78
	Mathematics	77,506	23	77
	Plain English Math	5,665		100
	Science	88,773	35	65
6	Reading	84,531	68	32
	Math	71,931	68	32
	Plain English Math	3,839	57	43
7	Reading	85,355	84	16
	Math	58,711	80	20
	Plain English Math	2,852	73	27

Grade	Subject	Total Number	Mode	
			Online (%)	Paper (%)
8	Reading	89,087	87	13
	Math	78,366	87	13
	Plain English Math	3,197	75	25
	Science	83,100	92	8
	Reading (Cumulative)	171		100
	Math (Cumulative)	125		100

Table 2.1.1.2 Percent of Tests Taken by Mode: Content-Specific History

Grade	Subject	Total Number	Mode	
			Online (%)	Paper (%)
CSH	Virginia Studies	87,644	33	67
	US History to 1877	74,578	70	30
	US History: 1877 to Present	72,573	91	9
	Civics and Economics	72,392	94	6

Table 2.1.1.3 Percent of Tests Taken by Mode: End-of-Course

Grade	Subject	Total Number	Mode	
			Online (%)	Paper (%)
EOC	English: Reading (1995)	24	8	92
	Earth Science	77,466	99	1
	Biology	92,676	99	1
	Chemistry	56,522	99	1
	Algebra I	97,959	99	1
	Geometry	88,064	99	1
	Algebra II (2001)	73	88	12
	Virginia & US History	79,634	99	1
	World History and Geography to 1500 A.D.	65,845	99	1
	World History and Geography: 1500 A.D. to Present	68,480	99	1
	World Geography	26,872	99	1
	English: Reading	76,901	98	2
	Algebra II (2001 Revised)	64,594	99	1
	Plain English Algebra I	1,628	97	3

2.1.2 Percent in Proficiency Level

The results in this section are based on all tests that were taken and scored with a valid score code. The tables below show the grade and subject area, the total number of tests taken, the percent passing at the Proficient and Advanced Proficient performance levels, and the overall passing rate. Tests taken on paper and online are combined in the calculation of the passing rates. For some subjects, the year of the standard is also included where tests are available from more than one set of standards.

Table 2.1.2.1 Grades 3-8 Passing Rates

Grade	Subject	N-Count	Proficiency Level		Overall Pass Rate (%)
			Proficient (%)	Advanced (%)	
3	Reading	87,035	50.6	34.9	85.5
	Math	90,008	43.8	44.5	88.3
	Science	88,480	46.6	41.9	88.5
	History	88,299	25.6	66.9	92.5
4	Reading	85,645	40.3	47.9	88.2
	Math	87,781	38.7	47.0	85.8
5	Reading	85,831	55.1	36.1	91.2
	Math	83,171	33.0	56.5	89.4
	Science	88,773	62.7	24.8	87.5
	Writing	84,927	59.5	27.1	86.6
6	Reading	84,531	49.7	36.0	85.7
	Math	75,770	41.3	30.2	71.5
7	Reading	85,355	43.2	44.9	88.2
	Math	61,563	43.8	25.1	68.9
8	Reading	89,087	44.9	40.7	85.6
	Math	81,563	33.0	49.5	82.5
	Science	83,100	50.1	40.3	90.3
	Writing	86,274	85.6	3.6	89.3
	Reading (Cumulative)	171	14.6	1.2	15.8
	Math (Cumulative)	125	6.4		6.4

Table 2.1.2.2 Content-Specific History Passing Rates

Subject	N-Count	Proficiency Level		Overall Pass Rate (%)
		Proficient (%)	Advanced (%)	
Virginia Studies	87,644	41.8	45.8	87.6
US History to 1877	74,578	48.1	25.7	73.8
US History: 1877 to present	72,573	37.6	54.3	91.9
Civics and Economics	72,392	51.0	32.8	83.8

Table 2.1.2.3 End-of-Course Passing Rates

Subject	N-Count	Proficiency Level		Overall Pass Rate (%)
		Proficient (%)	Advanced (%)	
Earth Science	77,466	59.6	18.4	78.0
Biology	92,676	62.8	16.1	78.9
Chemistry	56,522	71.4	14.6	86.0
Algebra I	99,587	59.7	27.6	87.4
Geometry	88,064	55.4	21.9	77.3
Virginia & US History	79,634	53.4	37.8	91.3
World History I[1]	65,845	49.4	37.8	87.2
World History II[2]	68,480	52.4	35.1	87.5
World Geography	26,872	50.1	26.4	76.5
English: Reading	76,901	42.9	46.9	89.8
English: Reading (1995)	24	54.2	8.3	62.5
Algebra II (Revised 2001)	64,594	59.5	24.0	83.6
Algebra II (2001)	73	53.4	8.2	61.6
Writing	76,547	55.8	33.9	89.7

[1]World History I refers to World History and Geography to 1500 A.D.
[2]World History II refers to World History and Geography: 1500 A.D. to Present

2.1.3 Raw Score Summary Statistics

Tables 2.1.3.1 through 2.1.3.7 show the raw score summary statistics for each newly constructed Virginia SOL multiple-choice test taken in the 2009 spring administration. Each table shows the grade and subject area of the test, number of examinees taking each test per core as well as the number of test items, observed raw score mean, median, standard deviation, and minimum and maximum values. Tables 2.1.3.1 through 2.1.3.3 represent the paper administration data, while tables 2.1.3.4 through 2.1.3.7 represent the online administration data.

Table 2.1.3.1 Summary Statistics for Grades 3-8 Reading, Mathematics and Plain English Mathematics Paper

Subject	Grade	Core	Items	N	Mean	Med	SD	Min	Max
Reading	3	1	35	44,017	28.8	30	5.2	1	35
		2	35	21,151	28.5	30	4.8	1	35
	4	1	35	41,394	30.0	32	5.0	1	35
		2	35	20,402	29.7	31	5.2	1	35
	5	1	40	40,101	34.1	36	5.3	1	40
		2	40	19,820	34.8	36	4.7	3	40
	6	1	45	18,741	37.5	39	6.5	4	45
		2	45	5,975	35.9	38	7.5	5	45
	7	1							
		2	45	6,724	34.6	37	8.0	4	45
	8	1							
		2	45	3,743	35.5	38	7.7	1	45
		3	45	2,840	34.9	37	7.1	5	45
Mathematics	3	1	50	42,345	43.6	45	6.0	1	50
		2	50	20,327	44.0	46	5.8	5	50
	4	1	50	39,967	40.3	43	7.9	1	50
		2	50	19,587	40.8	43	7.4	7	50
	5	1	50	34,439	43.2	45	6.4	5	50
		2	50	18,736	43.7	45	5.9	10	50
	6	1	50	15,748	39.6	42	7.9	3	50
		2	50	5,110	36.1	38	9.2	7	50
	7	1							
		2	50	6,361	34.8	37	9.8	1	50
	8	1							
		2	50	2,839	38.1	41	9.6	2	50
		3	50	2,065	37.7	40	9.1	4	50

Subject	Grade	Core	Items	N	Mean	Med	SD	Min	Max
Plain English Mathematics	3	2	50	5,982	36.7	39	9.1	1	50
	4	2	50	5,374	31.3	32	9.3	1	50
	5	2	50	5,056	35.4	37	9.1	1	50
	6	2	50	1,381	29.9	31	10.4	4	50
	7	2	50	372	23.2	22	9.1	6	50
	8	3	50	366	29.4	29	11.4	6	50

Table 2.1.3.2 Summary Statistics for Grades 3, 5 and 8 History and Science Paper

Subject	Grade	Core	Items	N	Mean	Med	SD	Min	Max
History	3	1	40	41,892	34.7	36	5.6	1	40
		2	40	20,436	35.2	37	5.1	5	40
Science	3	1	40	43,416	34.2	36	5.0	4	40
		2	40	20,064	33.9	35	5.3	1	40
	5	1	40	34,347	32.8	34	5.8	1	40
		2	40	17,914	33.4	35	5.5	1	40
	8	1							
		2	50	4,487	38.0	41	9.9	3	50

Table 2.1.3.3 Summary Statistics for Content-Specific History Paper

Subject	Core	Items	N	Mean	Med	SD	Min	Max
Virginia Studies	1	40	34,160	32.9	34	5.9	1	40
	2	40	18,857	33.1	35	6.0	3	40
US History to 1877	1	40	15,659	30.6	33	7.5	5	40
	2	40	4,309	28.8	31	7.9	5	40
US History: 1877 to Present	1							
	2	40	2,097	28.5	30	8.5	1	40
Civics and Economics	1							
	2	40	2,116	27.1	28	8.3	1	40

Table 2.1.3.4 Summary Statistics for Grades 3-8 Reading, Math, and Plain English Mathematics Online

Subject	Grade	Core	Items	N	Mean	Med	SD	Min	Max
Reading	3	1							
		2	35	10,373	27.8	29	5.2	2	35
	4	1							
		2	35	12,578	29.3	31	5.4	5	35
	5	1							
		2	40	14,630	34.2	36	5.0	1	40
	6	1	45	36,330	36.9	39	6.6	3	45
		2	45	11,934	37.1	39	6.4	5	45
	7	1	45	37,603	37.4	39	6.1	5	45
		2	45	24,245	37.6	39	6.3	1	45
	8	1	45	37,361	35.9	38	7.4	3	45
		2	45	19,129	36.7	39	6.6	4	45
		3	45	6,750	34.7	36	7.2	2	45
Mathematics	3	1							
		2	50	9,630	43.5	45	6.2	9	50
	4	1							
		2	50	11,419	39.8	42	7.7	9	50
	5	1							
		2	50	13,533	42.7	44	6.4	8	50
	6	1	50	29,514	38.1	40	8.0	4	50
		2	50	10,908	37.5	39	8.6	2	50
	7	1	50	24,491	33.6	35	8.3	5	50
		2	50	15,571	34.0	35	8.7	1	50
	8	1	50	31,438	41.6	44	7.2	4	50
		2	50	18,397	39.9	42	8.2	4	50
		3	50	5,186	38.5	41	8.4	6	50
Plain English Mathematics	3	2							
	4	2							
	5	2							
	6	2	50	1,199	28.0	28	9.8	4	49
	7	2	50	1,432	24.5	24	9.3	6	50
	8	3	50	1,531	30.4	32	11.0	3	50

Table 2.1.3.5 Summary Statistics for Grades 3, 5, and 8 History and Science Online

Subject	Grade	Core	Items	N	Mean	Med	SD	Min	Max
History	3	1							
		2	40	14,268	34.8	36	5.2	8	40
Science	3	1							
		2	40	13,212	34.4	36	5.0	6	40
	5	1							
		2	40	24,771	33.1	34	5.5	1	40
	8	1	50	50,590	40.2	42	7.6	4	50
		2	50	14,753	39.9	43	8.6	2	50

Table 2.1.3.6 Summary Statistics for Content-Specific History Online

Subject	Core	Items	N	Mean	Med	SD	Min	Max
Virginia Studies	1							
	2	40	23,017	33.1	35	6.0	5	40
US History to 1877	1	40	34,937	31.1	33	7.0	4	40
	2	40	8,681	31.6	33	6.8	4	40
US History: 1877 to Present	1	40	34,296	32.9	35	6.1	6	40
	2	40	21,640	33.4	35	6.1	5	40
Civics and Economics	1	40	36,350	30.7	32	7.0	5	40
	2	40	21,284	31.0	33	6.8	5	40

Table 2.1.3.7 Summary Statistics for High School End-of-Course Online

Subject	Core	Items	N	Mean	Med	SD	Min	Max
Earth Science	1	50	25,419	38.2	40	8.3	6	50
	2	50	20,400	36.5	38	8.7	4	50
	3	50	7,313	36.6	38	8.2	7	50
Biology	1	50	36,614	37.1	39	8.4	6	50
	2	50	24,092	35.0	37	9.4	3	50
	3	50	7,379	35.0	37	9.2	6	50
Chemistry	1	50	25,550	37.2	38	7.5	7	50
	2	50	12,618	36.1	37	7.7	8	50
	3	50	4,009	36.8	37	7.3	6	50
Algebra I	1	50	37,353	38.8	41	8.4	4	50
	2	50	26,253	36.6	38	8.6	4	50
	3	50	8,877	37.6	39	8.3	7	50
Geometry	1	45	33,082	36.2	38	6.9	5	45
	2	45	21,132	33.7	35	7.7	6	45
	3	45	6,977	33.5	34	7.5	7	45
Virginia & US History	1	60	32,191	47.2	49	8.6	8	60
	2	60	22,023	46.7	49	9.5	1	60
	3	60	6,924	46.8	49	9.8	10	60
World History and Geography to1500 A.D.	1	60	33,964	45.3	47	10.1	8	60
	2	60	10,590	42.7	44	10.7	4	60
	3	60	5,160	42.6	44	10.7	10	60
World History and Geography: 1500 A.D. to Present	1	60	29,346	45.8	48	9.9	8	60
	2	60	19,050	43.8	45	10.5	7	60
	3	60	5,325	43.8	46	10.3	2	60
World Geography	1	60	11,259	42.4	44	10.6	10	60
	2	60	7,018	44.6	46	9.3	9	60
English: Reading	1	50	31,802	41.1	43	6.6	8	50
	2	50	20,090	39.0	40	6.9	6	50
	3	50	6,736	40.1	42	7.3	9	50
Algebra II (Revised 2001)	1	50	27,326	40.4	42	7.6	6	50
	2	50	14,312	39.1	41	8.1	6	50
	3	50	4,570	39.3	41	7.9	7	50
Plain English Algebra I	3	50	789	32.1	33	9.4	6	50

Table 2.1.3.8 shows the raw score summary statistics for grades 5, 8, and End-of-Course Virginia *SOL* Writing tests taken in the 2009 spring administration. The table presents the number of examinees tested for every grade/core/prompt combination as well as the observed raw score mean, median, standard deviation, and minimum and maximum values. The maximum possible raw score for the Grade 5 Writing test is 44. It includes 20 multiple-choice questions and an essay item. The maximum possible raw score for the Grade 8 Writing test is 48. It includes 24 multiple-choice questions and an essay item. The maximum possible raw score for

the End-of-Course Writing test is 54. It includes 30 multiple-choice questions and an essay item. The Writing tests are administered only in the paper mode.

Table 2.1.3.8 Summary Statistics for Grades 5, 8 and EOC Writing Tests

Grade	Core	Prompt	N	Mean	Med	SD	Min	Max
5	1	5255	10,091	33.4	34	5.7	8	44
		5264	48,678	34.1	35	5.8	7	44
	2	5255	10,688	33.8	34	5.2	8	44
		5264	9,110	33.6	34	5.6	7	44
	3	5255	3,409	33.1	34	5.9	9	44
		5264	2,793	32.7	34	6.4	10	44
8	1	8257	49,100	37.7	38	6.0	7	48
		8259	10,112	36.6	37	6.2	8	48
	2	8257	9,394	37.3	38	6.3	9	48
		8259	11,101	37.3	38	5.9	9	48
	3	8257	2,674	37.3	38	6.5	9	48
		8259	3,195	37.5	38	6.3	11	48
End-of-Course	1	1656	8,984	44.4	46	6.1	10	54
		1663	40,442	44.9	46	6.0	12	54
	2	1656	9,664	44.3	45	5.8	10	54
		1663	8,337	44.5	46	6.1	14	54
	3	1656	2,102	44.4	46	6.7	14	54
		1663	2,161	43.2	45	6.7	11	54

2.2 Reliability Estimates for Multiple-Choice Assessments

2.2.1 Overall Reliability Estimates

This section addresses the overall reliability estimates for each newly constructed SOL test administered in spring 2009. Each table shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode combination (if a given core is **not** administered in a particular mode, then these cells are shaded). In all instances, the reliability coefficients are well-above the accepted lower limit of .70.

Table 2.2.1.1 shows the number of students used in the analyses and the results for Reading grades 3 though 8. For the online administrations, the Alphas ranged from .84 to .89, while in paper administrations, the Alphas ranged from .82 to .90.

Table 2.2.1.1 Cronbach's Alphas for Grades 3-8 Reading

Subject	Grade	Core	Online		Paper	
			N	Alpha	N	Alpha
Reading	3	1			44,017	0.85
		2	10,373	0.84	21,151	0.82
	4	1			41,394	0.86
		2	12,578	0.87	20,402	0.86
	5	1			40,101	0.86
		2	14,630	0.84	19,820	0.82
	6	1	36,330	0.88	18,741	0.88
		2	11,934	0.87	5,975	0.90
	7	1	37,603	0.86		
		2	24,245	0.87	6,724	0.90
	8	1	37,361	0.89		
		2	19,129	0.88	3,743	0.90
		3	6,750	0.88	2,840	0.88

Table 2.2.1.2 shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode combination for Mathematics grades 3 through 8. For the online administrations, the Alphas ranged from .87 to .91. The Alphas ranged from .87 to .93 for the paper administrations.

Table 2.2.1.2 Cronbach's Alphas for Grades 3-8 Mathematics

Subject	Grade	Core	Online		Paper	
			N	Alpha	N	Alpha
Math	3	1			42,345	0.87
		2	9,630	0.88	20,327	0.87
	4	1			39,967	0.90
		2	11,419	0.89	19,587	0.89
	5	1			34,439	0.88
		2	13,533	0.88	18,736	0.87
	6	1	29,514	0.89	15,748	0.90
		2	10,908	0.90	5,110	0.91
	7	1	24,491	0.87		
		2	15,571	0.89	6,361	0.92
	8	1	31,438	0.89		
		2	18,397	0.91	2,839	0.93
		3	5,186	0.90	2,065	0.92

Table 2.2.1.3 shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode combination for Plain English Mathematics grades 3 through 8. For both the online and paper administrations, the Alphas ranged from .88 to .93.

Table 2.2.1.3 Cronbach's Alphas for Grades 3-8 Plain English Mathematics

Subject	Grade	Core	Online		Paper	
			N	Alpha	N	Alpha
Plain English Math	3	2			5,982	0.91
	4	2			5,374	0.90
	5	2			5,056	0.90
	6	2	1,199	0.90	1,381	0.91
	7	2	1,432	0.88	372	0.88
	8	3	1,531	0.93	366	0.93

Table 2.2.1.4 shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode combination for History and Science grades 3, 5, and 8. In History, for the online administrations, the Alpha was .86. For the paper administrations, the Alphas ranged from .87 to .88. In Science, for the online administrations, the Alphas ranged from .85 to .91. For the paper administrations, the Alphas ranged from .84 to .93.

Table 2.2.1.4 Cronbach's Alphas for Grades 3, 5, and 8 History and Science

Subject	Grade	Core	Online		Paper	
			N	Alpha	N	Alpha
History	3	1			41,892	0.88
		2	14,268	0.86	20,436	0.87
Science	3	1			43,416	0.84
		2	13,212	0.85	20,064	0.86
	5	1			34,347	0.86
		2	24,771	0.85	17,914	0.85
	8	1	50,590	0.89		
		2	14,753	0.91	4,487	0.93

Table 2.2.1.5 shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode combination for Content-Specific History tests. For the online administrations, the Alphas ranged from .87 to .89. The Alphas ranged from .87 to .91 for the paper administrations.

Table 2.2.1.5 Cronbach's Alphas for Content-Specific History Tests

Subject	Core	Online		Paper	
		N	Alpha	N	Alpha
Virginia Studies	1			34,160	0.87
	2	23,017	0.87	18,857	0.87
United States History to 1877	1	34,937	0.89	15,659	0.90
	2	8,681	0.89	4,309	0.90
United States History from 1877 to Present	1	34,296	0.87		
	2	21,640	0.88	2,097	0.91
Civics and Economics	1	36,350	0.88		
	2	21,284	0.88	2,116	0.90

Table 2.2.1.6 shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode combination for End-of-Course tests. The Alphas ranged from .85 to .92 for the online administrations. There were no paper administrations for the newly constructed End-of-Course tests in the spring 2009 administration.

Table 2.2.1.6 Cronbach's Alphas for High School End-of-Course Tests

Subject	Core	Online	
		N	Alpha
Earth Science	1	25,419	0.90
	2	20,400	0.90
	3	7,313	0.88
Biology	1	36,614	0.89
	2	24,092	0.90
	3	7,379	0.90
Chemistry	1	25,550	0.86
	2	12,618	0.87
	3	4,009	0.85
Algebra I	1	37,353	0.91
	2	26,253	0.89
	3	8,877	0.89
Geometry	1	33,082	0.88
	2	21,132	0.88
	3	6,977	0.88
Virginia & United States History	1	32,191	0.89
	2	22,023	0.91
	3	6,924	0.92
World History I	1	33,964	0.91
	2	10,590	0.91
	3	5,160	0.92
World History II	1	29,346	0.91
	2	19,050	0.91
	3	5,325	0.91
World Geography	1	11,259	0.91
	2	7,018	0.89
English: Reading/Lit. & Res.	1	31,802	0.86
	2	20,090	0.85
	3	6,736	0.88
Algebra II	1	27,326	0.89
	2	14,312	0.89
	3	4,570	0.89
Plain English Algebra I	3	789	0.90

2.2.2 Reliability Estimates by Gender

Tables 2.2.2.1 through 2.2.2.6 address the subgroup reliability results by gender for SOL tests administered in spring 2009. Each table shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode/gender combination. In all instances, the reliability coefficients are well-above the accepted lower limit of .70. Students not reporting their gender are excluded from these results.

Table 2.2.2.1 shows the results for Reading grades 3 though 8. In the online administrations, the Alphas for the females ranged from .83 to .89, while the Alphas for males ranged from .84 to .90. In the paper administrations, the Alphas for the females ranged from .80 to .89, while the Alphas for the males ranged from .83 to .91.

Table 2.2.2.1 Cronbach's Alphas for Grades 3-8 Reading by Gender

Subject	Grade	Core	Online				Paper			
			Female		Male		Female		Male	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
Reading	3	1					21,650	0.84	22,367	0.86
		2	5,121	0.83	5,252	0.84	10,473	0.81	10,678	0.83
	4	1					20,429	0.85	20,965	0.86
		2	6,223	0.86	6,355	0.87	10,119	0.85	10,283	0.87
	5	1					19,608	0.85	20,493	0.87
		2	7,215	0.83	7,415	0.84	9,767	0.80	10,053	0.83
	6	1	18,096	0.87	18,234	0.88	9,169	0.86	9,572	0.89
		2	5,863	0.86	6,071	0.87	2,944	0.88	3,031	0.91
	7	1	18,741	0.85	18,862	0.87				
		2	12,045	0.86	12,200	0.88	3,247	0.89	3,477	0.91
	8	1	18,324	0.89	19,037	0.90				
		2	9,416	0.87	9,713	0.88	1,861	0.89	1,882	0.91
		3	3,396	0.87	3,354	0.88	1,365	0.85	1,475	0.89

Table 2.2.2.2 shows the results for Mathematics grades 3 though 8. In the online administrations, the Alphas for the females ranged from .87 to .90, while the Alphas for males ranged from .88 to .92. In the paper administrations, the Alphas for the females ranged from .86 to .91, while the Alphas for the males ranged from .87 to .94.

Table 2.2.2.2 Cronbach's Alphas for Grades 3-8 Mathematics by Gender

Subject	Grade	Core	Online				Paper			
			Female		Male		Female		Male	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
Math	3	1					21,134	0.87	21,211	0.88
		2	4,811	0.87	4,819	0.88	10,170	0.87	10,157	0.88
	4	1					20,046	0.90	19,921	0.90
		2	5,653	0.89	5,766	0.90	9,853	0.89	9,734	0.90
	5	1					17,223	0.87	17,216	0.89
		2	6,729	0.87	6,804	0.88	9,410	0.86	9,326	0.87
	6	1	15,011	0.88	14,503	0.89	7,731	0.89	8,017	0.90
		2	5,455	0.90	5,453	0.90	2,535	0.90	2,575	0.91
	7	1	12,291	0.87	12,200	0.88				
		2	7,812	0.88	7,759	0.89	3,234	0.91	3,127	0.92
	8	1	15,532	0.88	15,906	0.90				
		2	9,080	0.90	9,317	0.92	1,381	0.91	1,458	0.94
		3	2,562	0.90	2,624	0.91	982	0.90	1,083	0.93

Table 2.2.2.3 shows the results for Plain English Mathematics grades 3 though 8. In the online administrations, the Alphas for the females ranged from .89 to .92, while the Alphas for males ranged from .88 to .93. In the paper administrations, the Alphas for the females ranged from .87 to .93, while the Alphas for the males ranged from .88 to .93.

Table 2.2.2.3 Cronbach's Alphas for Grades 3-8 Plain English Mathematics by Gender

Subject	Grade	Core	Online				Paper			
			Female		Male		Female		Male	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
Plain English Math	3	2					2,296	0.90	3,686	0.91
	4	2					2,060	0.90	3,314	0.90
	5	2					1,805	0.90	3,251	0.90
	6	2	439	0.90	760	0.90	539	0.91	842	0.92
	7	2	573	0.89	859	0.88	132	0.87	240	0.88
	8	3	601	0.92	930	0.93	123	0.93	243	0.93

Table 2.2.2.4 shows the results for History and Science grades 3, 5, and 8. In History administered on paper, the Alphas for the females ranged from .86 to .88, while the Alphas for the males ranged from .87 to .89. In History administered online, Alpha was .86 for the females, and .87 for the males. In Science administered on paper, the Alphas for the females ranged from .83 to .92, while the Alphas for the males ranged from .84 to .94. In Science administered online, the Alphas for the females ranged from .84 to .91, while the Alphas for the males ranged from .85 to .92.

Table 2.2.2.4 Cronbach's Alphas for Grades 3, 5, and 8 History and Science by Gender

Subject	Grade	Core	Online				Paper			
			Female		Male		Female		Male	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
History	3	1					20,510	0.88	21,382	0.89
		2	7,037	0.86	7,231	0.87	10,024	0.86	10,412	0.87
Science	3	1					21,285	0.83	22,131	0.84
		2	6,499	0.85	6,713	0.85	9,798	0.85	10,266	0.86
	5	1					16,606	0.85	17,741	0.87
		2	12,142	0.84	12,629	0.85	8,811	0.84	9,103	0.86
	8	1	24,982	0.89	25,608	0.90				
		2	7,325	0.91	7,428	0.92	2,123	0.92	2,364	0.94

Table 2.2.2.5 shows the results for the Content-Specific History tests. For the online administrations, the Alphas ranged from .87 to .88 for the females, while the Alphas ranged from .87 to .89 for the males. For the paper administrations, the Alphas ranged from .85 to .90 for the females and from .88 to .92 for the males.

Table 2.2.2.5 Cronbach's Alphas for Content-Specific History Tests by Gender

Subject	Core	Online				Paper			
		Female		Male		Female		Male	
		N	Alpha	N	Alpha	N	Alpha	N	Alpha
Virginia Studies	1					16,775	0.85	17,385	0.88
	2	11,251	0.87	11,766	0.88	9,289	0.86	9,568	0.88
United States History to 1877	1	17,540	0.88	17,397	0.89	7,573	0.90	8,086	0.90
	2	4,289	0.88	4,392	0.89	2,069	0.89	2,240	0.90
United States History from 1877 to Present	1	17,075	0.87	17,221	0.87				
	2	10,647	0.88	10,993	0.88	971	0.90	1,126	0.92
Civics and Economics	1	18,172	0.88	18,178	0.89				
	2	10,391	0.88	10,893	0.89	1,027	0.89	1,089	0.91

Table 2.2.2.6 shows the results for high school End-of Course tests. For the online administrations, the Alphas ranged from .84 to .91 for the females, while the Alphas ranged from .86 to .92 for the males. There were no paper administrations for the newly constructed End-of-Course tests in the spring 2009 administration.

Table 2.2.2.6 Cronbach's Alphas for High School End-of-Course Tests by Gender

Subject	Core	Online			
		Female		Male	
		N	Alpha	N	Alpha
Earth Science	1	12,636	0.89	12,783	0.90
	2	10,307	0.89	10,093	0.90
	3	3,598	0.87	3,715	0.89
Biology	1	18,126	0.88	18,488	0.90
	2	12,271	0.90	11,821	0.91
	3	3,727	0.90	3,652	0.90
Chemistry	1	13,077	0.86	12,473	0.87
	2	6,881	0.86	5,737	0.87
	3	2,180	0.84	1,829	0.86
Algebra I	1	18,509	0.90	18,844	0.91
	2	13,222	0.89	13,031	0.90
	3	4,463	0.89	4,414	0.90
Geometry	1	16,615	0.89	16,467	0.88
	2	10,863	0.88	10,269	0.89
	3	3,629	0.88	3,348	0.88
Virginia & United States History	1	15,958	0.88	16,233	0.89
	2	11,258	0.90	10,765	0.91
	3	3,561	0.91	3,363	0.92
World History I	1	16,726	0.91	17,238	0.92
	2	5,335	0.91	5,255	0.92
	3	2,561	0.91	2,599	0.92
World History II	1	14,597	0.90	14,749	0.91
	2	9,756	0.91	9,294	0.92
	3	2,747	0.91	2,578	0.92
World Geography	1	5,800	0.91	5,459	0.91
	2	3,646	0.89	3,372	0.90
English: Reading/Lit. & Res.	1	15,938	0.86	15,864	0.86
	2	10,305	0.85	9,785	0.86
	3	3,488	0.88	3,248	0.88
Algebra II	1	13,970	0.88	13,356	0.89
	2	7,779	0.89	6,533	0.90
	3	2,460	0.88	2,110	0.90
Plain English Algebra I	3	331	0.90	458	0.90

2.2.3 Reliability Estimates by Ethnic Group

Tables 2.2.3.1 through 2.2.3.6 address the subgroup reliability results by ethnic group for SOL tests administered in spring 2009. The student population is distributed in such a way that analyses were only possible for two ethnic groups: black and white. Each table shows the number of students used in the analyses and the associated Cronbach's Alpha for each grade/core/mode/ethnic combination. In all instances, the reliability coefficients are well above

the accepted lower limit of .70. Students not reporting their ethnicity are excluded from these results.

Table 2.2.3.1 shows the results for Reading grades 3 though 8. In the online administrations, the Alphas for the black students ranged from .84 to .89, while the Alphas for the white students ranged from .82 to .88. In the paper administrations, the Alphas for the black students ranged from .82 to .89, while the Alphas for the white students ranged from .80 to .89.

Table 2.2.3.1 Cronbach's Alphas for Grades 3-8 Reading by Ethnic Group

Subject	Grade	Core	Online				Paper			
			Black		White		Black		White	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
Reading	3	1					13,151	0.84	22,338	0.83
		2	1,725	0.84	7,531	0.83	4,687	0.82	11,750	0.80
	4	1					12,459	0.85	20,820	0.84
		2	2,151	0.87	9,113	0.86	4,598	0.86	11,277	0.85
	5	1					12,182	0.85	20,147	0.84
		2	2,748	0.84	10,280	0.82	4,409	0.83	11,008	0.80
	6	1	9,578	0.87	21,587	0.86	4,148	0.88	9,470	0.86
		2	2,310	0.87	7,378	0.86	2,141	0.89	2,890	0.88
	7	1	10,475	0.86	22,282	0.84				
		2	3,542	0.88	13,891	0.86	3,194	0.89	2,931	0.89
	8	1	7,632	0.89	21,545	0.88				
		2	5,751	0.87	11,026	0.87	1,553	0.89	1,873	0.89
		3	2,948	0.87	3,086	0.85	900	0.89	1,264	0.85

Table 2.2.3.2 shows the results for Mathematics grades 3 though 8. In the online administrations, the Alphas for the black students ranged from .86 to .90, while the Alphas for the white students ranged from .86 to .91. In the paper administrations, the Alphas for the black students ranged from .87 to .93, while the Alphas for the white students ranged from .85 to .93.

Table 2.2.3.2 Cronbach's Alphas for Grades 3-8 Mathematics by Ethnic Group

Subject	Grade	Core	Online				Paper			
			Black		White		Black		White	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
Math	3	1					12,332	0.87	21,602	0.85
		2	1,676	0.88	6,976	0.87	4,436	0.88	11,328	0.85
	4	1					11,697	0.89	20,287	0.88
		2	1,968	0.90	8,331	0.88	4,337	0.89	10,908	0.87
	5	1					10,950	0.88	17,096	0.87
		2	2,664	0.88	9,367	0.86	4,095	0.87	10,446	0.85
	6	1	8,443	0.88	16,957	0.88	3,618	0.89	8,085	0.87
		2	2,233	0.90	6,202	0.89	1,751	0.90	2,542	0.91
	7	1	7,052	0.86	14,073	0.86				
		2	2,661	0.87	8,664	0.88	2,517	0.89	2,827	0.91
	8	1	7,016	0.89	17,645	0.88				
		2	5,524	0.90	10,586	0.91	1,485	0.93	1,105	0.93
		3	2,197	0.90	2,465	0.88	634	0.91	1,037	0.91

Table 2.2.3.3 shows the results for Plain English Mathematics grades 3 though 8. In the online administrations, the Alphas for the black students ranged from .85 to .91, while the Alphas for the white students ranged from .87 to .92. In the paper administrations, the Alphas for the black students ranged from .88 to .94, while the Alphas for the white students ranged from .86 to .91.

Table 2.2.3.3 Cronbach's Alphas for Grades 3-8 Plain English Mathematics by Ethnic Group

Subject	Grade	Core	Online				Paper			
			Black		White		Black		White	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
Plain English Math	3	2					1,462	0.91	1,969	0.91
	4	2					1,455	0.89	1,943	0.89
	5	2					1,487	0.89	1,901	0.90
	6	2	316	0.88	551	0.89	313	0.90	509	0.91
	7	2	286	0.85	528	0.87	184	0.88	129	0.86
	8	3	473	0.91	545	0.92	128	0.94	153	0.91

Table 2.2.3.4 shows the results for History and Science grades 3, 5 and 8. In History administered on paper, Alpha was .88 for the black students, while the Alphas for the white students ranged from .84 to .87. In History administered online, Alpha was .86 for both the black and white students. In Science administered on paper, the Alphas for the black students ranged from .83 to .91, while the Alphas for the white students ranged from .80 to .92. In Science administered online, the Alphas for the black students ranged from .83 to .90, while the Alphas for the white students ranged from .83 to .89.

Table 2.2.3.4 Cronbach's Alphas for Grades 3, 5, and 8 History and Science by Ethnic Group

Subject	Grade	Core	Online				Paper			
			Black		White		Black		White	
			N	Alpha	N	Alpha	N	Alpha	N	Alpha
History	3	1					12,565	0.88	20,697	0.87
		2	2,675	0.86	10,080	0.86	4,507	0.88	11,057	0.84
Science	3	1					12,265	0.83	23,525	0.80
		2	2,346	0.85	9,470	0.84	5,312	0.85	8,918	0.84
	5	1					10,446	0.84	15,939	0.82
		2	5,226	0.83	16,705	0.83	4,215	0.85	9,341	0.82
	8	1	11,691	0.88	29,275	0.87				
		2	4,197	0.90	8,236	0.89	1,823	0.91	1,707	0.92

Table 2.2.3.5 shows the results for Content-Specific History tests. In the online administrations, the Alphas for the black students ranged from .87 to .88, while the Alphas for the white students ranged from .86 to .88. In the paper administrations, the Alphas for the black students ranged from .85 to .90, while the Alphas for the white students ranged from .85 to .93.

Table 2.2.3.5 Cronbach's Alphas for Content-Specific History Tests by Ethnic Group

Subject	Core	Online				Paper			
		Black		White		Black		White	
		N	Alpha	N	Alpha	N	Alpha	N	Alpha
Virginia Studies	1					10,144	0.85	16,332	0.85
	2	5,074	0.87	15,300	0.86	4,155	0.87	10,164	0.86
United States History to 1877	1	9,291	0.88	20,765	0.88	3,232	0.89	7,370	0.88
	2	1,238	0.88	6,524	0.88	2,053	0.88	1,853	0.90
United States History from 1877 to Present	1	9,256	0.87	20,245	0.86				
	2	2,896	0.87	12,975	0.86	1,335	0.90	587	0.93
Civics and Economics	1	10,921	0.87	20,911	0.87				
	2	2,848	0.87	12,940	0.87	1,348	0.89	606	0.91

Table 2.2.3.6 shows the results for high school End-of-Course tests. In the online administrations, the Alphas for the black students ranged from .82 to .91, while the Alphas for the white students ranged from .84 to .91. There were no paper administrations for the newly constructed End-of-Course tests in the spring 2009 administration.

Table 2.2.3.6 Cronbach's Alphas for High School End-of-Course Tests by Ethnic Group

Subject	Core	Online			
		Black		White	
		N	Alpha	N	Alpha
Earth Science	1	6,464	0.88	14,707	0.86
	2	7,859	0.87	9,437	0.88
	3	1,770	0.87	5,114	0.87
Biology	1	7,086	0.86	20,738	0.87
	2	8,824	0.88	11,434	0.89
	3	1,944	0.87	5,029	0.90
Chemistry	1	3,574	0.83	15,590	0.85
	2	3,362	0.84	7,514	0.86
	3	725	0.82	3,065	0.85
Algebra I	1	7,195	0.89	21,042	0.90
	2	9,197	0.88	12,921	0.89
	3	2,316	0.89	6,003	0.89
Geometry	1	5,893	0.86	19,656	0.86
	2	7,134	0.86	11,020	0.87
	3	1,827	0.84	4,760	0.87
Virginia & United States History	1	5,994	0.88	19,089	0.87
	2	7,806	0.90	11,194	0.90
	3	1,633	0.91	4,918	0.91
World History I	1	7,255	0.90	19,036	0.90
	2	3,767	0.89	5,875	0.91
	3	1,638	0.90	3,234	0.91
World History II	1	4,880	0.90	16,957	0.90
	2	5,922	0.90	9,762	0.91
	3	1,127	0.90	3,878	0.91
World Geography	1	3,516	0.89	6,478	0.90
	2	1,714	0.89	3,931	0.88
English: Reading/Lit. & Res.	1	5,746	0.85	19,210	0.84
	2	6,809	0.83	10,756	0.84
	3	1,490	0.88	4,875	0.86
Algebra II	1	3,941	0.87	17,046	0.88
	2	4,216	0.88	8,240	0.89
	3	961	0.88	3,355	0.88
Plain English Algebra I	3	181	0.87	227	0.88

2.3 Reliability Estimates for Writing Assessments

2.3.1 Stratified Alpha

The tables below present reliability results in Writing for grades 5, 8, and End-of-Course, overall as well as by gender and ethnic group, for each combination of multiple-choice core and writing prompt for the Main and Alternate administration.

Table 2.3.1.1 shows the number of students used in the analyses and the reliability results for each combination of multiple-choice core and writing prompt for the Main and Alternate administration for grades 5, 8, and End-of-Course Writing. For grade 5, the Stratified Alphas ranged from .85 to .89. For grade 8, the Stratified Alphas ranged from .88 to .90. For End-of-Course, the Stratified Alphas ranged from .89 to .92.

Table 2.3.1.1 Stratified Alphas for Grades 5, 8, and End-of-Course Tests

Grade	Core	Prompt	N	Alpha
Writing 5	1	5255	10,091	0.88
		5264	48,678	0.89
	2	5255	10,688	0.85
		5264	9,110	0.87
	3	5255	3,409	0.88
		5264	2,793	0.89
Writing 8	1	8257	49,100	0.88
		8259	10,112	0.89
	2	8257	9,394	0.89
		8259	11,101	0.89
	3	8257	2,674	0.90
		8259	3,195	0.90
Writing End-of-Course	1	1656	8,984	0.91
		1663	40,442	0.90
	2	1656	9,664	0.89
		1663	8,337	0.90
	3	1656	2,102	0.92
		1663	2,161	0.91

Table 2.3.1.2 shows the reliability results for grades 5, 8, and End-of-Course Writing by gender for each combination of multiple-choice core and writing prompt for the Main and Alternate administration. For grade 5, the Stratified Alphas for the females ranged from .84 to .89, while the Stratified Alphas for the males ranged from .85 to .89. For grade 8, the Stratified Alphas for the females ranged from .87 to .90, while the Stratified Alphas for the males ranged from .88 to .90. For End-of Course, the Stratified Alphas for the females ranged from .88 to .91, while the Stratified Alphas for the males ranged from .90 to .92.

Table 2.3.1.2 Stratified Alphas for Grades 5, 8 and End-of-Course Tests by Gender

Grade	Core	Prompt	Female		Male	
			N	Alpha	N	Alpha
Writing 5	1	5255	4,982	0.88	5,109	0.88
		5264	23,959	0.88	24,719	0.88
	2	5255	5,290	0.84	5,398	0.85
		5264	4,476	0.86	4,634	0.86
	3	5255	1,721	0.86	1,688	0.88
		5264	1,409	0.89	1,384	0.89
Writing 8	1	8257	24,210	0.87	24,890	0.88
		8259	4,973	0.87	5,139	0.89
	2	8257	4,733	0.89	4,661	0.89
		8259	5,515	0.88	5,586	0.89
	3	8257	1,291	0.89	1,383	0.90
		8259	1,622	0.90	1,573	0.89
Writing End-of-Course	1	1656	4,563	0.90	4,421	0.91
		1663	20,525	0.90	19,917	0.91
	2	1656	4,846	0.88	4,818	0.90
		1663	4,096	0.89	4,241	0.90
	3	1656	1,063	0.91	1,039	0.92
		1663	1,124	0.90	1,037	0.91

Table 2.3.1.3 shows the reliability results for grades 5, 8, and End-of-Course Writing by ethnic group for each combination of multiple-choice core and writing prompt for the Main and Alternate administration. For grade 5, the Stratified Alphas for the black students ranged from .83 to .88, while the Stratified Alphas for white students ranged from .84 to .89. For grade 8, the Stratified Alphas for the black students ranged from .86 to .90, while the Stratified Alphas for white students ranged from .87 to .90. For End-of-Course, the Stratified Alphas for the black students ranged from .88 to .90, while the Stratified Alphas for white students ranged from .89 to .90.

Table 2.3.1.3 Stratified Alphas for Grades 5, 8, and End-of-Course Writing by Ethnic Group

Level	Core	Prompt	Black		White	
			N	Alpha	N	Alpha
Writing 5	1	5255	2,242	0.87	6,202	0.87
		5264	11,760	0.87	27,798	0.88
	2	5255	2,895	0.83	6,568	0.84
		5264	2,088	0.84	5,816	0.87
	3	5255	1,674	0.86	1,363	0.88
		5264	753	0.88	1,590	0.89
Writing 8	1	8257	11,869	0.86	28,436	0.87
		8259	2,482	0.88	6,142	0.88
	2	8257	2,070	0.88	6,091	0.89
		8259	2,941	0.88	6,983	0.88
	3	8257	650	0.88	1,616	0.90
		8259	1,458	0.90	1,410	0.89
Writing End-of-Course	1	1656	2,138	0.90	5,567	0.90
		1663	9,558	0.89	23,896	0.90
	2	1656	2,336	0.88	6,420	0.89
		1663	1,728	0.88	5,620	0.90
	3	1656	636	0.90	1,235	0.90
		1663	515	0.89	1,375	0.90

2.3.2 Inter-Rater Reliability

There were two new writing prompts administered in the spring 2009 administration in grades 5, 8, and End-of-Course Writing. The following section addresses inter-rater reliability of the scoring process. Tables 2.3.2.1 through 2.3.2.3 address the grade 5, grade 8, and End-of-Course prompts, respectively.

Table 2.3.2.1 shows that across both grade 5 writing prompts the percent of scores that are *at least* adjacent (perfect agreement plus adjacent agreement) ranges from 98% to 100%.

Table 2.3.2.1 Inter-Rater Reliability for Grade 5 Writing Assessment: Prompts 5255 and 5264

Prompt	Trait	N	Perfect Agree (%)	Adjacent (%)	Non -Adjacent (%)
5264	Composing	64,918	72	28	1
	Written Expression	64,918	71	28	1
	Usage and Mechanics	64,918	67	31	2
5255	Composing	24,419	72	27	0
	Written Expression	24,419	72	28	1
	Usage and Mechanics	24,419	67	32	1

Table 2.3.2.2 shows that across both grade 8 writing prompts the percent of scores that are *at least* adjacent (perfect agreement plus adjacent agreement) ranges from 99% to 100%.

Table 2.3.2.2 Inter-Rater Reliability for Grade 8 Writing Assessment: Prompts 8257 and 8259

Prompt	Trait	N	Perfect Agree (%)	Adjacent (%)	Non -Adjacent (%)
8257	Composing	65,605	70	29	1
	Written Expression	65,605	71	29	1
	Usage and Mechanics	65,605	67	32	1
8259	Composing	24,789	73	26	1
	Written Expression	24,789	73	27	1
	Usage and Mechanics	24,789	67	32	1

Table 2.3.2.3 shows that across both End-of-Course writing prompts the percent of scores that are *at least* adjacent (perfect agreement plus adjacent agreement) ranges from 99% to 100%

Table 2.3.2.3 Inter-Rater Reliability for EOC Writing Assessment: Prompts 1656 and 1663

Prompt	Trait	N	Perfect Agree (%)	Adjacent (%)	Non -Adjacent (%)
1663	Composing	56,807	73	26	0
	Written Expression	56,807	73	26	0
	Usage and Mechanics	56,807	67	32	1
1656	Composing	21,280	76	24	0
	Written Expression	21,280	75	25	0
	Usage and Mechanics	21,280	67	32	1

2.4 Decision Consistency and Accuracy Indices

Tables 2.4.1 through 2.4.7 present the number of examinees taking each Virginia SOL multiple-choice test in the 2009 spring administration and the proportion of accurate classifications, false positives, false negatives, and consistent classifications for passing (proficient) on the test. Tables 2.4.1-2.4.3 present the paper administration data, while tables 2.4.4-2.4.7 present the online administration data (if a given core is **not** administered in a particular mode, then these cells are shaded). Table 2.4.8 presents the results for grades 5, 8, and End-of-Course Writing tests.

Table 2.4.1 shows the results for grades 3 through 8 multiple-choice paper tests in reading and mathematics. Decision accuracy ranged from 94% to 97% in Reading, from 92% to 97% in Mathematics, and from 91% to 93% in Plain English Mathematics across grades and cores. Decision consistency ranged from 91% to 95% in Reading, from 89% to 96% in Mathematics, and from 87% to 91% in Plain English Mathematics across grades and cores.

Table 2.4.1 Decision Consistency and Accuracy Indices for Grades 3-8 Reading, Mathematics, and Plain English Mathematics

Subject	Grade	Core	N	Accuracy	False Positive	False Negative	Consistency
Reading	3	1	44,017	0.94	0.03	0.03	0.92
		2	21,151	0.94	0.03	0.03	0.92
	4	1	41,394	0.96	0.02	0.03	0.94
		2	20,402	0.95	0.02	0.03	0.93
	5	1	40,101	0.97	0.01	0.02	0.95
		2	19,820	0.97	0.01	0.02	0.95
	6	1	18,741	0.96	0.02	0.02	0.94
		2	5,975	0.94	0.03	0.03	0.91
	7	1					
		2	6,724	0.94	0.03	0.03	0.92
	8	1					
		2	3,743	0.94	0.03	0.03	0.92
		3	2,840	0.95	0.02	0.03	0.93
Math	3	1	42,345	0.96	0.02	0.02	0.94
		2	20,327	0.97	0.02	0.02	0.95
	4	1	39,967	0.96	0.02	0.02	0.94
		2	19,587	0.96	0.02	0.02	0.94
	5	1	34,439	0.97	0.01	0.02	0.95
		2	18,736	0.97	0.01	0.02	0.96
	6	1	15,748	0.94	0.03	0.03	0.92
		2	5,110	0.92	0.04	0.04	0.89
	7	1					
		2	6,361	0.93	0.03	0.04	0.90
	8	1					
		2	2,839	0.95	0.03	0.03	0.92
		3	2,065	0.94	0.03	0.03	0.91
Plain English Math	3	2	5,982	0.92	0.04	0.04	0.89
	4	2	5,374	0.91	0.05	0.05	0.87
	5	2	5,056	0.91	0.04	0.04	0.88
	6	2	1,381	0.91	0.05	0.04	0.88
	7	2	372	0.93	0.04	0.03	0.91
	8	3	366	0.93	0.04	0.03	0.90

Table 2.4.2 shows the results for the grades 3, 5, and 8 paper tests in History and Science. Decision accuracy was 97% in History and ranged from 94% to 96% in Science. Decision consistency ranged from 95% to 96% in History and from 92% to 94% in Science across grades and cores.

Table 2.4.2 Decision Consistency and Accuracy Indices for Grades 3, 5 and 8 History and Science

Subject	Grade	Core	N	Accuracy	False Positive	False Negative	Consistency
History	3	1	41,892	0.97	0.01	0.02	0.95
		2	20,436	0.97	0.01	0.02	0.96
Science	3	1	43,416	0.95	0.02	0.03	0.93
		2	20,064	0.95	0.02	0.03	0.93
	5	1	34,347	0.94	0.02	0.03	0.92
		2	17,914	0.95	0.02	0.03	0.93
	8	1					
		2	4,487	0.96	0.02	0.02	0.94

Table 2.4.3 shows the results for the paper version of the Content-Specific History tests. Decision accuracy ranged from 92% to 96%, and decision consistency ranged from 89% to 94% across different subject areas and cores.

Table 2.4.3 Decision Consistency and Accuracy Indices for Content-Specific History

Subject	Core	N	Accuracy	False Positive	False Negative	Consistency
Virginia Studies	1	34,160	0.95	0.02	0.03	0.93
	2	18,857	0.96	0.02	0.03	0.94
US History to 1877	1	15,659	0.93	0.03	0.04	0.90
	2	4,309	0.92	0.04	0.04	0.89
US History: from 1877	1					
	2	2,097	0.93	0.04	0.04	0.90
Civics and Economics	1					
	2	2,116	0.93	0.03	0.04	0.90

Table 2.4.4 shows the results for grades 3 through 8 multiple-choice online tests in Reading, Mathematics, and Plain English Mathematics. Decision accuracy ranged from 93% to 96% in Reading, from 91% to 97% in Mathematics, and from 92% to 93% in Plain English Mathematics across grades and cores. Decision consistency ranged from 90% to 94% in Reading, from 88% to 96% in Mathematics, and from 88% to 90% in Plain English Mathematics across grades and cores.

Table 2.4.4 Decision Consistency and Accuracy Indices for Grades 3-8 Reading, Math, and Plain English Mathematics Online Tests

Subject	Grade	Core	N	Accuracy	False Positive	False Negative	Consistency
Reading	3	1					
		2	10,373	0.93	0.03	0.04	0.90
	4	1					
		2	12,578	0.95	0.02	0.03	0.93
	5	1					
		2	14,630	0.95	0.02	0.03	0.93
	6	1	36,330	0.95	0.02	0.03	0.93
		2	11,934	0.95	0.02	0.03	0.93
	7	1	37,603	0.96	0.02	0.02	0.94
		2	24,245	0.96	0.02	0.02	0.94
	8	1	37,361	0.96	0.02	0.02	0.94
		2	19,129	0.96	0.02	0.02	0.94
		3	6,750	0.94	0.03	0.03	0.92
Math	3	1					
		2	9,630	0.97	0.01	0.02	0.96
	4	1					
		2	11,419	0.95	0.02	0.03	0.93
	5	1					
		2	13,533	0.95	0.02	0.03	0.93
	6	1	29,514	0.91	0.04	0.04	0.88
		2	10,908	0.92	0.04	0.04	0.89
	7	1	24,491	0.91	0.04	0.05	0.88
		2	15,571	0.91	0.04	0.04	0.88
	8	1	31,438	0.96	0.02	0.02	0.94
		2	18,397	0.95	0.02	0.03	0.93
		3	5,186	0.94	0.03	0.03	0.92
Plain English Math	3	2					
	4	2					
	5	2					
	6	2	1,199	0.92	0.05	0.03	0.89
	7	2	1,432	0.92	0.05	0.04	0.88
	8	3	1,531	0.93	0.04	0.03	0.90

Table 2.4.5 shows the results for grades 3, 5, and 8 online version tests in History and Science. Decision accuracy ranged from 95% to 97% in Science and was 97% in History. Decision consistency was 96% in History and ranged from 92% to 96% in Science across grades and cores.

Table 2.4.5 Decision Consistency and Accuracy Indices for Grades 3, 5, and 8 History and Science Online Tests

Subject	Grade	Core	N	Accuracy	False Positive	False Negative	Consistency
History	3	1					
		2	14,268	0.97	0.01	0.02	0.96
Science	3	1					
		2	13,212	0.96	0.02	0.02	0.94
	5	1					
		2	24,771	0.95	0.02	0.03	0.92
	8	1	50,590	0.97	0.01	0.02	0.96
		2	14,753	0.96	0.02	0.02	0.95

Table 2.4.6 shows the results for the online versions of the Content-Specific History tests. Decision accuracy ranged from 92% to 97%, and decision consistency ranged from 89% to 96% across different subject areas and cores.

Table 2.4.6 Decision Consistency and Accuracy Indices for Content-Specific History Online Tests

Subject	Core	N	Accuracy	False Positive	False Negative	Consistency
Virginia Studies	1					
	2	23,017	0.94	0.02	0.03	0.92
US History to 1877	1	34,937	0.92	0.04	0.04	0.89
	2	8,681	0.93	0.03	0.04	0.91
US History: from 1877	1	34,296	0.97	0.01	0.02	0.96
	2	21,640	0.97	0.01	0.02	0.96
Civics and Economics	1	36,350	0.94	0.03	0.03	0.91
	2	21,284	0.95	0.02	0.03	0.93

Table 2.4.7 shows the results for the online versions of the high school End-of-Course tests. Decision accuracy ranged from 92% to 98%, and decision consistency ranged from 89% to 97% across different subject areas and cores.

Table 2.4.7 Decision Consistency and Accuracy Indices for High School End-of-Course Online Tests

Subject	Core	N	Accuracy	False Positive	False Negative	Consistency
Earth Science	1	25,419	0.95	0.02	0.03	0.93
	2	20,400	0.93	0.03	0.04	0.91
	3	7,313	0.94	0.03	0.03	0.91
Biology	1	36,614	0.95	0.02	0.03	0.94
	2	24,092	0.94	0.03	0.03	0.91
	3	7,379	0.93	0.03	0.04	0.91
Chemistry	1	25,550	0.95	0.02	0.03	0.93
	2	12,618	0.94	0.02	0.04	0.91
	3	4,009	0.96	0.02	0.03	0.94
Algebra I	1	37,353	0.97	0.01	0.02	0.96
	2	26,253	0.96	0.02	0.02	0.94
	3	8,877	0.96	0.01	0.02	0.95
Geometry	1	33,082	0.95	0.02	0.03	0.93
	2	21,132	0.92	0.03	0.04	0.89
	3	6,977	0.93	0.03	0.04	0.90
Virginia & United States History	1	32,191	0.98	0.01	0.01	0.97
	2	22,023	0.97	0.01	0.02	0.96
	3	6,924	0.97	0.01	0.02	0.96
World History I	1	33,964	0.97	0.01	0.02	0.96
	2	10,590	0.95	0.02	0.03	0.93
	3	5,160	0.96	0.02	0.02	0.95
World History II	1	29,346	0.97	0.01	0.02	0.96
	2	19,050	0.95	0.02	0.03	0.93
	3	5,325	0.95	0.02	0.03	0.93
World Geography	1	11,259	0.94	0.03	0.03	0.91
	2	7,018	0.94	0.02	0.03	0.92
English: Reading/Lit. & Res.	1	31,802	0.97	0.01	0.01	0.96
	2	20,090	0.96	0.02	0.02	0.94
	3	6,736	0.97	0.01	0.02	0.95
Algebra II	1	27,326	0.96	0.02	0.02	0.94
	2	14,312	0.94	0.03	0.03	0.91
	3	4,570	0.94	0.03	0.03	0.91
Plain English Algebra I	3	789	0.93	0.03	0.04	0.90

Table 2.4.8 shows the results for grades 5, 8, and End-of-Course Writing tests. The Writing tests were administered only in the paper mode. Decision accuracy ranged from 93% to 97%. Decision consistency ranged from 90% to 96%.

Table 2.4.8 Decision Consistency and Accuracy Indices for Grades 5, 8, and End-of-Course Writing Tests

Grade	Core	Prompt	N	Accuracy	False Positive	False Negative	Consistency
5	1	5255	10,091	0.94	0.03	0.03	0.92
		5264	48,678	0.95	0.02	0.03	0.92
	2	5255	10,688	0.94	0.03	0.03	0.92
		5264	9,110	0.94	0.03	0.03	0.91
	3	5255	3,409	0.94	0.03	0.03	0.91
		5264	2,793	0.93	0.03	0.04	0.90
8	1	8257	49,100	0.96	0.02	0.02	0.94
		8259	10,112	0.95	0.02	0.03	0.93
	2	8257	9,394	0.95	0.02	0.03	0.92
		8259	11,101	0.96	0.02	0.02	0.94
	3	8257	2,674	0.95	0.03	0.03	0.92
		8259	3,195	0.96	0.02	0.02	0.94
End-of-Course	1	1656	8,984	0.97	0.01	0.02	0.96
		1663	40,442	0.97	0.01	0.02	0.96
	2	1656	9,664	0.97	0.01	0.02	0.96
		1663	8,337	0.97	0.01	0.02	0.96
	3	1656	2,102	0.97	0.01	0.02	0.96
		1663	2,161	0.96	0.02	0.02	0.94

2.5 Raw Score to Scale Score (RSSS) Conversion Tables and Conditional SEM

Table 2.5.1 RSSS Conversions for Grade 3 Reading

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	99	65	93	66			95	66
2	146	47	141	48			142	48
3	176	40	170	40			172	40
4	197	35	192	35			194	35
5	215	32	210	32			212	32
6	230	30	225	30			227	30
7	243	28	239	29			240	29
8	255	27	251	27			253	27
9	266	26	262	26			264	26
10	276	25	273	26			275	26
11	286	25	283	25			285	25
12	295	24	292	25			295	25
13	304	24	302	24			304	24
14	313	24	311	24			313	24
15	321	23	320	24			322	24
16	330	23	329	24			331	24
17	338	23	337	24			340	24
18	347	23	346	24			349	24
19	355	23	355	24			357	24
20	364	23	364	24			366	24
21	372	24	373	24			375	24
22	381	24	382	25			385	25
23	390	24	392	25			395	25
24	400	25	402	26			405	26
25	409	25	413	26			415	26
26	420	26	424	27			426	27
27	431	27	436	28			438	28
28	443	28	449	30			452	30
29	456	30	464	31			466	31
30	471	32	480	34			483	34
31	489	35	499	37			502	37
32	511	40	523	41			525	41
33	540	47	554	49			556	49
34	587	65	600				600	
35	600		600				600	

Table 2.5.2 RSSS Conversions for Grade 3 Mathematics

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	72	63	82	63			96	63
2	117	45	127	45			140	45
3	145	38	154	37			168	37
4	164	33	173	33			187	33
5	180	30	189	30			203	30
6	194	28	202	28			217	28
7	205	26	214	26			228	26
8	216	25	224	25			239	25
9	225	24	234	24			248	24
10	234	23	242	23			257	23
11	242	22	250	22			265	22
12	250	22	258	21			272	21
13	257	21	265	21			279	21
14	264	21	272	20			286	20
15	271	20	278	20			293	20
16	278	20	285	20			299	20
17	284	20	291	19			306	19
18	290	19	297	19			312	19
19	296	19	303	19			318	19
20	302	19	308	19			323	19
21	308	19	314	19			329	19
22	314	19	320	19			335	19
23	319	19	325	18			340	19
24	325	19	331	18			346	19
25	331	19	336	18			351	19
26	336	19	342	18			357	19
27	342	19	347	18			363	19
28	348	19	353	19			368	19
29	354	19	358	19			374	19
30	359	19	364	19			380	19
31	365	19	370	19			386	19
32	371	20	375	19			392	19
33	378	20	381	19			398	20
34	384	20	388	20			404	20
35	391	20	394	20			410	20
36	397	21	401	20			417	21
37	404	21	408	21			424	21
38	412	22	415	21			432	22
39	420	22	422	22			439	22
40	428	23	431	23			448	23
41	437	24	439	24			456	24
42	447	25	449	25			466	25
43	457	26	459	26			477	26

44	469	28	471	28		489	28
45	483	30	485	30		502	30
46	499	33	501	33		519	33
47	519	38	521	38		539	38
48	546	45	548	45		567	46
49	592	63	594	63		600	
50	600		600			600	

Table 2.5.3 RSSS Conversions for Grade 3 Plain English Mathematics

Raw Score	Paper		Online	
	Core 2		Core 2	
	SS	SEM	SS	SEM
0	0			
1	86	63		
2	131	45		
3	158	37		
4	177	33		
5	193	30		
6	207	28		
7	218	26		
8	229	25		
9	238	24		
10	247	23		
11	255	22		
12	262	21		
13	269	21		
14	276	20		
15	283	20		
16	289	20		
17	296	19		
18	302	19		
19	307	19		
20	313	19		
21	319	19		
22	325	19		
23	330	19		
24	336	19		
25	341	19		
26	347	19		
27	352	19		
28	358	19		
29	364	19		
30	369	19		
31	375	19		
32	381	19		
33	387	19		
34	393	20		
35	400	20		
36	406	21		
37	413	21		
38	421	22		
39	428	22		
40	436	23		
41	445	24		
42	455	25		
43	465	26		

44	477	28	
45	491	30	
46	507	33	
47	527	38	
48	554	46	
49	600		
50	600		

Table 2.5.4 RSSS Conversions for Grade 3 Science

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	80	62	104	61			99	61
2	125	45	148	44			143	44
3	153	38	175	37			170	37
4	174	33	194	32			190	33
5	191	31	210	30			206	30
6	206	29	224	28			220	28
7	218	27	236	26			232	26
8	230	26	247	25			243	25
9	241	25	257	24			253	24
10	250	24	266	23			262	23
11	260	23	275	23			271	23
12	269	23	283	22			279	22
13	277	22	291	22			287	22
14	285	22	299	21			295	21
15	293	22	306	21			303	21
16	301	21	314	21			310	21
17	309	21	321	21			317	21
18	316	21	328	21			324	21
19	323	21	335	21			331	21
20	331	21	342	21			339	21
21	338	21	349	21			346	21
22	345	21	356	21			353	21
23	353	21	364	21			360	21
24	360	21	371	21			368	21
25	368	21	379	21			375	21
26	376	22	387	22			383	22
27	384	22	395	22			391	22
28	392	22	403	23			400	23
29	401	23	412	23			408	23
30	410	24	421	24			418	24
31	419	24	431	25			428	25
32	429	25	442	26			439	26
33	440	26	454	27			450	27
34	453	28	467	29			464	29
35	467	30	482	31			479	31
36	483	33	500	34			496	34
37	503	37	522	39			518	38
38	531	44	551	46			547	46
39	575	61	598	63			594	63
40	600		600				600	

Table 2.5.5 RSSS Conversions for Grade 3 History and Social Studies

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	61	76	49	76			55	76
2	116	55	104	55			110	55
3	149	45	137	46			144	46
4	173	40	162	41			169	41
5	192	36	182	37			189	37
6	209	34	199	34			206	35
7	223	32	214	33			221	33
8	236	30	228	31			235	31
9	248	29	240	30			247	30
10	259	28	252	29			259	29
11	270	27	263	28			270	28
12	280	27	273	27			280	28
13	289	26	283	27			290	27
14	298	26	293	27			299	27
15	307	26	302	26			309	26
16	316	25	311	26			318	26
17	324	25	320	26			327	26
18	332	25	329	26			336	26
19	341	25	337	25			344	26
20	349	25	346	25			353	25
21	357	25	355	25			362	26
22	366	25	363	26			370	26
23	374	25	372	26			379	26
24	383	25	381	26			388	26
25	392	26	390	26			397	26
26	401	26	399	26			407	27
27	410	27	409	27			416	27
28	420	27	419	27			426	28
29	430	28	429	28			437	28
30	440	29	440	29			448	29
31	452	30	451	30			459	30
32	464	31	464	31			472	31
33	478	33	477	33			485	33
34	493	35	492	34			500	35
35	510	37	509	37			518	37
36	530	41	529	41			538	41
37	555	46	554	46			563	46
38	589	55	588	55			597	55
39	600		600				600	
40	600		600				600	

Table 2.5.6 RSSS Conversions for Grade 4 Reading

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	111	64	119	63			113	63
2	157	46	164	46			158	46
3	186	39	192	38			186	38
4	207	34	213	34			207	34
5	225	31	229	31			224	31
6	240	29	243	29			238	29
7	253	28	256	27			251	27
8	264	26	267	26			262	26
9	275	25	278	25			273	25
10	285	24	287	24			283	24
11	295	24	297	24			292	24
12	304	23	305	23			301	23
13	312	23	314	23			310	23
14	320	23	322	22			318	23
15	329	22	330	22			326	22
16	337	22	338	22			334	22
17	344	22	346	22			342	22
18	352	22	354	22			350	22
19	360	22	362	22			358	22
20	368	22	370	22			366	22
21	376	22	378	23			375	23
22	384	23	387	23			383	23
23	393	23	395	23			392	23
24	401	23	404	24			401	24
25	410	24	413	24			410	24
26	420	25	423	25			420	25
27	430	26	434	26			430	26
28	441	27	445	27			442	27
29	454	28	458	29			455	29
30	468	30	472	31			469	31
31	484	33	489	34			486	34
32	504	38	510	38			507	38
33	532	45	538	46			535	46
34	577	63	584	63			580	63
35	600		600				600	

Table 2.5.7 RSSS Conversions for Grade 4 Mathematics

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	81	69	80	69			71	69
2	130	49	129	49			120	49
3	160	41	159	41			150	41
4	182	36	180	36			172	36
5	199	33	198	33			190	33
6	214	30	213	30			204	31
7	227	29	225	29			217	29
8	238	27	237	27			229	27
9	249	26	247	26			240	26
10	259	25	257	25			249	25
11	268	24	266	24			259	24
12	276	24	275	24			267	24
13	284	23	283	23			275	23
14	292	23	291	23			283	23
15	299	22	298	22			291	22
16	306	22	305	22			298	22
17	313	21	312	22			305	22
18	320	21	319	21			312	22
19	327	21	326	21			319	21
20	333	21	333	21			326	21
21	340	21	339	21			332	21
22	346	21	346	21			339	21
23	352	20	352	21			345	21
24	358	20	359	21			352	21
25	364	20	365	21			358	21
26	371	20	371	21			365	21
27	377	20	378	21			371	21
28	383	20	384	21			378	21
29	389	21	391	21			384	21
30	395	21	398	21			391	21
31	402	21	404	21			398	22
32	408	21	411	22			405	22
33	415	21	418	22			412	22
34	422	22	426	22			419	22
35	429	22	433	23			427	23
36	436	22	441	23			435	23
37	444	23	449	24			443	24
38	451	23	457	24			451	24
39	460	24	466	25			460	25
40	469	25	476	26			470	26
41	478	26	486	27			480	27
42	488	27	497	28			491	28
43	500	28	509	29			503	29

44	512	30	522	31		517	31
45	527	33	538	33		532	34
46	544	36	556	37		551	37
47	565	41	578	42		574	42
48	595	49	600			600	
49	600		600			600	
50	600		600			600	

Table 2.5.8 RSSS Conversions for Grade 4 Plain English Mathematics

Raw Score	Paper		Online	
	Core 2		Core 2	
	SS	SEM	SS	SEM
0	0			
1	85	69		
2	135	49		
3	164	41		
4	186	36		
5	203	33		
6	218	30		
7	231	29		
8	242	27		
9	253	26		
10	262	25		
11	271	24		
12	280	24		
13	288	23		
14	296	23		
15	303	22		
16	310	22		
17	317	22		
18	324	21		
19	331	21		
20	337	21		
21	344	21		
22	350	21		
23	357	21		
24	363	21		
25	369	21		
26	376	21		
27	382	21		
28	389	21		
29	395	21		
30	402	21		
31	408	21		
32	415	22		
33	422	22		
34	430	22		
35	437	23		
36	445	23		
37	453	24		
38	461	24		
39	470	25		
40	479	26		
41	489	27		
42	500	28		
43	512	29		

44	526	31	
45	541	33	
46	559	37	
47	581	41	
48	600		
49	600		
50	600		

Table 2.5.9 RSSS Conversions for Grade 5 Reading

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	157	53	138	54			127	54
2	196	38	178	39			168	40
3	219	32	202	33			194	33
4	237	28	221	29			213	29
5	251	26	235	26			228	27
6	263	24	248	24			241	25
7	273	23	259	23			252	23
8	283	22	269	22			262	22
9	292	21	277	21			271	21
10	300	20	286	20			280	20
11	307	20	293	20			287	20
12	315	19	301	19			295	19
13	322	19	308	19			302	19
14	328	18	314	18			309	19
15	335	18	321	18			315	18
16	341	18	327	18			322	18
17	347	18	333	18			328	18
18	354	18	340	18			334	18
19	360	18	346	18			340	18
20	366	18	352	18			346	18
21	372	18	358	18			352	18
22	378	18	364	18			358	18
23	384	18	370	18			364	18
24	391	18	376	18			371	18
25	397	18	382	18			377	18
26	404	19	389	18			383	18
27	410	19	395	19			390	19
28	418	19	402	19			397	19
29	425	20	409	20			404	19
30	433	20	417	20			411	20
31	441	21	425	21			420	21
32	450	22	434	22			428	22
33	460	23	443	23			438	23
34	470	24	454	24			448	24
35	483	26	466	26			460	26
36	497	29	480	28			474	28
37	515	32	497	32			491	32
38	538	38	521	38			515	38
39	577	53	559	53			553	53
40	600		600				600	

Table 2.5.10 RSSS Conversions for Grade 5 Mathematics

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	6	80	15	79			0	
2	64	57	72	57			56	57
3	98	48	106	47			91	47
4	124	42	131	41			116	42
5	144	38	151	38			135	38
6	161	35	168	35			152	35
7	176	33	182	33			167	33
8	189	31	195	31			180	31
9	201	30	207	30			192	30
10	213	29	218	29			203	29
11	223	28	228	28			213	28
12	233	27	238	27			222	27
13	242	27	247	26			232	26
14	251	26	256	26			240	26
15	259	26	264	25			249	25
16	268	25	272	25			257	25
17	276	25	280	25			265	25
18	283	24	287	24			272	24
19	291	24	295	24			280	24
20	298	24	302	24			287	24
21	306	24	310	24			294	24
22	313	24	317	24			301	24
23	320	24	324	24			309	24
24	327	24	331	24			316	24
25	334	24	338	24			323	24
26	341	24	345	24			330	24
27	348	24	352	24			337	24
28	355	24	360	24			344	24
29	363	24	367	24			351	24
30	370	24	374	24			359	24
31	377	24	382	24			366	24
32	385	24	389	25			374	25
33	393	25	397	25			382	25
34	400	25	405	25			390	25
35	409	25	414	26			398	26
36	417	26	422	26			407	26
37	426	26	431	27			415	27
38	435	27	441	28			425	27
39	445	28	451	28			435	28
40	455	29	461	29			445	29
41	466	30	473	30			457	30
42	478	31	485	32			469	32
43	491	33	498	33			482	33

44	506	35	514	36		498	36
45	523	38	531	38		515	38
46	543	42	552	42		536	42
47	568	47	577	48		561	48
48	600		600			596	58
49	600		600			600	
50	600		600			600	

Table 2.5.11 RSSS Conversions for Grade 5 Plain English Mathematics

Raw Score	Paper		Online	
	Core 2		Core 2	
	SS	SEM	SS	SEM
0	0			
1	19	79		
2	75	57		
3	109	47		
4	134	41		
5	154	38		
6	171	35		
7	186	33		
8	199	31		
9	210	30		
10	221	29		
11	231	28		
12	241	27		
13	250	26		
14	258	26		
15	267	25		
16	275	25		
17	282	24		
18	290	24		
19	297	24		
20	305	24		
21	312	24		
22	319	23		
23	326	23		
24	333	23		
25	340	23		
26	347	23		
27	354	23		
28	361	23		
29	368	24		
30	375	24		
31	382	24		
32	390	24		
33	397	25		
34	405	25		
35	413	25		
36	421	26		
37	430	26		
38	439	27		
39	449	28		
40	459	29		
41	470	30		
42	482	31		
43	496	33		

44	510	35	
45	527	38	
46	548	42	
47	573	48	
48	600		
49	600		
50	600		

Table 2.5.12 RSSS Conversions for Grade 5 Science

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	168	50	165	50			167	50
2	204	36	202	36			203	36
3	226	30	224	30			225	30
4	242	26	240	27			242	27
5	255	24	253	24			255	24
6	266	22	264	23			266	23
7	276	21	274	21			276	21
8	284	20	283	20			285	20
9	292	19	291	20			293	20
10	300	19	299	19			301	19
11	307	18	306	18			308	19
12	313	18	313	18			315	18
13	320	17	320	18			322	18
14	326	17	326	17			328	17
15	332	17	332	17			334	17
16	337	17	338	17			340	17
17	343	17	344	17			346	17
18	349	16	350	17			352	17
19	354	16	356	17			358	17
20	360	16	361	17			363	17
21	365	16	367	17			369	17
22	371	17	373	17			375	17
23	376	17	379	17			381	17
24	382	17	384	17			387	17
25	388	17	390	17			393	17
26	394	17	396	17			399	17
27	400	17	403	18			405	18
28	407	18	409	18			412	18
29	413	18	416	18			418	18
30	420	19	423	19			426	19
31	428	19	431	19			433	20
32	436	20	439	20			441	20
33	445	21	448	21			450	21
34	454	23	457	22			460	23
35	466	24	468	24			471	24
36	479	26	481	26			485	27
37	495	30	498	30			501	30
38	517	36	519	36			523	36
39	553	50	555	50			559	50
40	600		600				600	

Table 2.5.13 RSSS Conversions for Grade 6 Reading

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0		0		0	
1	96	63	100	62	103	62	98	63
2	141	45	144	45	148	45	144	45
3	169	38	172	37	176	38	171	38
4	189	33	191	33	196	33	192	33
5	206	30	207	30	212	30	209	30
6	220	28	221	28	226	28	222	28
7	232	26	233	26	238	26	235	26
8	243	25	243	25	249	25	246	25
9	253	24	253	24	259	24	255	24
10	262	23	262	23	268	23	264	23
11	271	23	270	22	277	23	273	22
12	279	22	278	22	285	22	281	22
13	286	21	285	21	293	22	289	21
14	294	21	292	21	300	21	296	21
15	301	21	299	20	308	21	303	20
16	308	20	306	20	315	20	309	20
17	315	20	312	20	321	20	316	20
18	321	20	318	19	328	20	322	20
19	328	20	325	19	334	20	329	19
20	334	20	331	19	341	20	335	19
21	340	20	337	19	347	20	341	19
22	347	20	343	19	354	20	347	19
23	353	20	349	19	360	20	353	19
24	359	20	355	19	366	20	359	19
25	366	20	361	19	372	20	365	19
26	372	20	367	19	379	20	371	19
27	378	20	373	19	385	20	378	20
28	385	20	379	20	392	20	384	20
29	392	20	385	20	399	20	390	20
30	399	21	392	20	406	21	397	20
31	406	21	399	21	413	21	404	21
32	413	21	406	21	420	21	411	21
33	421	22	413	21	428	22	418	21
34	429	22	421	22	436	22	426	22
35	437	23	429	23	444	23	434	23
36	446	24	438	24	453	24	443	24
37	456	25	447	25	463	25	452	25
38	467	26	458	26	474	26	463	26
39	479	28	469	27	486	28	474	28
40	492	30	483	30	499	30	488	30
41	508	33	498	33	515	33	504	33
42	528	37	518	37	535	37	523	37
43	556	45	545	45	563	45	550	45

44	600		589	62	600		595	62
45	600		600		600		600	

Table 2.5.14 RSSS Conversions for Grade 6 Mathematics

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0		0		0	
1	23	77	25	77	12	77	11	77
2	78	55	80	55	67	55	66	55
3	111	46	113	46	100	46	100	46
4	135	40	138	40	125	40	125	41
5	155	37	157	37	144	37	144	37
6	171	34	174	34	161	34	161	34
7	186	32	189	32	175	32	176	32
8	198	30	201	30	188	30	189	31
9	210	29	213	29	200	29	201	29
10	221	28	224	28	211	28	212	28
11	231	27	234	27	221	27	222	27
12	241	26	244	26	230	27	232	27
13	250	26	253	26	240	26	241	26
14	258	25	261	25	248	25	250	26
15	267	25	270	25	257	25	258	25
16	275	25	278	24	265	25	267	25
17	283	24	286	24	273	24	275	24
18	290	24	293	24	280	24	282	24
19	298	24	301	24	288	24	290	24
20	305	24	308	23	295	24	297	24
21	313	23	315	23	303	23	305	23
22	320	23	322	23	310	23	312	23
23	327	23	329	23	317	23	319	23
24	334	23	336	23	324	23	326	23
25	341	23	343	23	331	23	333	23
26	348	23	350	23	338	23	341	23
27	356	23	357	23	346	23	348	23
28	363	23	365	23	353	23	355	23
29	370	23	372	23	360	23	362	23
30	377	24	379	23	367	24	369	24
31	385	24	386	24	375	24	377	24
32	392	24	394	24	382	24	384	24
33	400	24	401	24	390	24	392	24
34	408	25	409	24	398	25	400	25
35	416	25	417	25	406	25	408	25
36	425	25	426	25	414	25	417	25
37	433	26	434	26	423	26	425	26
38	443	27	443	26	432	27	435	27
39	452	27	453	27	442	27	444	27
40	463	28	463	28	452	28	454	28
41	474	29	474	29	463	29	465	29
42	486	31	486	30	474	30	477	31
43	499	32	498	32	487	32	490	32

44	513	34	513	34	502	34	505	34
45	530	37	529	37	518	37	521	37
46	550	41	549	40	538	40	541	40
47	574	46	573	46	562	46	565	46
48	600		600		595	55	599	55
49	600		600		600		600	
50	600		600		600		600	

Table 2.5.15 RSSS Conversions for Grade 6 Plain English Mathematics

Raw Score	Paper		Online	
	Core 2		Core 2	
	SS	SEM	SS	SEM
0	0		0	
1	34	77	17	77
2	89	55	72	55
3	122	46	105	46
4	146	40	130	40
5	165	36	149	37
6	181	34	166	34
7	195	32	180	32
8	208	30	193	30
9	219	29	204	29
10	230	28	215	28
11	239	27	225	27
12	249	26	235	26
13	257	25	244	26
14	266	25	252	25
15	274	24	260	25
16	282	24	268	24
17	289	24	276	24
18	297	23	283	24
19	304	23	291	23
20	311	23	298	23
21	318	23	305	23
22	325	23	312	23
23	332	23	319	23
24	338	23	326	23
25	345	23	332	23
26	352	23	339	23
27	359	23	346	23
28	366	23	353	23
29	373	23	360	23
30	380	23	367	23
31	387	23	374	23
32	394	24	382	24
33	402	24	389	24
34	409	24	397	24
35	417	25	405	25
36	425	25	413	25
37	434	26	421	26
38	443	26	430	26
39	452	27	440	27
40	462	28	450	28
41	473	29	460	29
42	485	30	472	30
43	497	32	485	32

44	512	34	499	34
45	528	37	515	37
46	548	40	535	40
47	572	46	559	46
48	600		592	55
49	600		600	
50	600		600	

Table 2.5.16 RSSS Conversions for Grade 7 Reading

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0			0		0		0	
1			97	66	73	67	84	66
2			145	48	122	48	132	48
3			173	39	152	40	161	40
4			194	35	174	36	182	35
5			211	32	192	33	199	32
6			225	29	207	30	213	30
7			238	28	220	29	226	28
8			249	26	232	27	237	27
9			259	25	243	26	248	25
10			269	24	254	25	257	25
11			277	24	263	25	266	24
12			286	23	272	24	275	23
13			294	22	281	23	283	23
14			301	22	289	23	291	22
15			309	22	297	23	298	22
16			316	21	305	22	306	22
17			323	21	312	22	313	21
18			330	21	320	22	320	21
19			336	21	327	22	327	21
20			343	21	334	21	334	21
21			350	21	341	21	340	21
22			356	21	348	21	347	21
23			363	21	355	21	354	21
24			369	21	362	21	361	21
25			376	21	369	21	368	21
26			383	21	376	21	374	21
27			390	21	383	22	381	21
28			396	21	390	22	388	22
29			404	22	397	22	396	22
30			411	22	405	22	403	22
31			418	22	412	23	411	23
32			426	23	420	23	419	23
33			434	23	429	23	427	23
34			443	24	437	24	436	24
35			452	25	446	25	445	25
36			461	25	456	26	455	26
37			472	27	467	27	465	27
38			483	28	478	28	477	28
39			496	30	491	30	490	30
40			510	32	505	32	504	32
41			527	35	523	35	521	35
42			548	40	544	40	543	40
43			577	48	573	48	572	48

44		600		600		600	
45		600		600		600	

Table 2.5.17 RSSS Conversions for Grade 7 Mathematics

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0			0		0		0	
1			24	80	34	81	27	80
2			81	58	92	58	85	58
3			116	48	128	48	120	48
4			141	42	154	43	146	42
5			162	38	175	39	167	39
6			179	36	193	36	184	36
7			194	33	208	34	200	34
8			207	32	222	32	213	32
9			220	30	235	31	226	31
10			231	29	246	30	238	30
11			242	28	257	29	248	29
12			252	28	267	28	259	28
13			261	27	277	27	268	27
14			270	26	286	27	278	27
15			279	26	295	26	287	26
16			287	26	304	26	295	26
17			296	25	312	25	304	26
18			304	25	320	25	312	25
19			311	25	328	25	320	25
20			319	24	336	25	328	25
21			327	24	344	24	336	25
22			334	24	351	24	343	24
23			341	24	359	24	351	24
24			349	24	366	24	358	24
25			356	24	374	24	366	24
26			363	24	381	24	373	24
27			371	24	388	24	381	24
28			378	24	396	24	388	24
29			385	24	403	24	396	25
30			393	24	411	25	403	25
31			401	25	419	25	411	25
32			408	25	427	25	419	25
33			416	25	435	25	427	25
34			424	25	443	26	435	26
35			433	26	451	26	444	26
36			441	26	460	27	453	26
37			450	27	469	27	462	27
38			460	28	479	28	471	28
39			469	28	489	29	481	28
40			480	29	500	30	492	29
41			491	30	511	31	503	30
42			503	32	524	32	515	32
43			517	33	537	34	529	33

44		532	35	553	36	544	35
45		549	38	570	39	561	38
46		569	42	591	42	581	42
47		594	48	600		600	
48		600		600		600	
49		600		600		600	
50		600		600		600	

Table 2.5.18 RSSS Conversions for Grade 7 Plain English Mathematics

Raw Score	Paper		Online	
	Core 2		Core 2	
	SS	SEM	SS	SEM
0	0		0	
1	30	80	41	80
2	87	58	98	58
3	122	48	133	48
4	147	42	158	42
5	167	38	179	38
6	184	35	196	35
7	199	33	211	33
8	212	32	224	32
9	224	30	236	30
10	236	29	248	29
11	246	28	258	28
12	256	27	268	27
13	265	27	277	27
14	274	26	286	26
15	282	26	295	26
16	290	25	303	25
17	298	25	311	25
18	306	25	319	25
19	314	24	327	24
20	321	24	334	24
21	328	24	341	24
22	336	24	349	24
23	343	24	356	24
24	350	24	363	24
25	357	24	370	24
26	364	24	377	24
27	371	24	385	24
28	378	24	392	24
29	385	24	399	24
30	393	24	406	24
31	400	24	414	24
32	408	25	421	25
33	415	25	429	25
34	423	25	437	25
35	431	26	445	26
36	440	26	454	26
37	449	27	463	27
38	458	27	472	27
39	467	28	482	28
40	478	29	492	29
41	489	30	503	30
42	501	31	515	31
43	514	33	528	33

44	529	35	543	35
45	546	38	560	38
46	566	42	580	42
47	591	48	600	
48	600		600	
49	600		600	
50	600		600	

Table 2.5.19 RSSS Conversions for Grade 8 Reading

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0			0		0		0		0		0	
1			96	63	106	63	121	63	94	63	108	63
2			141	45	151	45	167	46	140	46	153	45
3			169	38	179	38	195	38	168	38	181	38
4			189	33	199	33	215	33	188	34	201	33
5			206	30	215	30	232	30	205	31	218	30
6			219	28	229	28	246	28	219	29	231	28
7			232	27	241	27	258	27	231	27	244	27
8			242	25	252	25	269	25	242	26	254	25
9			252	24	262	24	279	24	253	25	264	24
10			261	23	271	23	288	23	262	24	273	23
11			270	23	280	23	297	23	271	23	282	23
12			278	22	288	22	305	22	279	22	290	22
13			286	22	296	22	313	22	287	22	298	22
14			293	21	303	21	320	21	295	22	305	21
15			300	21	310	21	327	21	302	21	312	21
16			307	21	317	21	334	21	309	21	319	21
17			314	20	324	20	341	20	316	21	326	20
18			321	20	330	20	348	20	323	20	333	20
19			327	20	337	20	354	20	330	20	339	20
20			334	20	343	20	361	20	337	20	346	20
21			340	20	350	20	367	20	343	20	352	20
22			346	20	356	20	374	20	350	20	358	20
23			352	20	362	20	380	20	356	20	365	20
24			359	20	368	20	386	20	363	20	371	20
25			365	20	375	20	393	20	369	20	377	20
26			371	20	381	20	399	20	376	20	384	20
27			378	20	388	20	406	20	382	20	390	20
28			385	20	394	20	412	20	389	21	397	20
29			391	21	401	21	419	21	396	21	404	21
30			398	21	408	21	426	21	403	21	411	21
31			405	21	415	21	433	21	410	21	418	21
32			413	22	423	22	440	22	418	22	425	22
33			420	22	431	22	448	22	426	22	433	22
34			429	23	439	23	456	23	434	23	441	23
35			437	23	448	24	465	23	443	24	450	24
36			446	24	457	24	474	24	452	25	459	24
37			457	25	467	26	484	25	462	26	469	25
38			468	27	478	27	494	26	473	27	480	27
39			480	28	491	29	506	28	486	29	493	28
40			494	31	505	31	520	30	500	31	507	31
41			511	34	521	34	536	33	517	34	523	34
42			531	38	542	38	556	38	537	38	544	38
43			559	46	570	46	584	45	566	46	572	46

44		600		600		600		600		600	
45		600		600		600		600		600	

Table 2.5.20 RSSS Conversions for Grade 8 Mathematics

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0			0		0		0		0		0	
1			3	81	6	81	1	81	0		5	81
2			60	58	64	58	59	58	51	58	63	58
3			95	48	99	48	94	48	87	49	98	48
4			121	42	124	42	120	43	113	43	123	42
5			141	38	144	38	141	39	133	39	144	39
6			158	36	161	35	158	36	151	36	161	36
7			173	33	176	33	173	34	167	34	176	34
8			186	32	189	32	187	32	180	32	190	32
9			198	30	201	30	199	31	193	31	202	31
10			210	29	212	29	211	30	205	30	213	29
11			220	28	223	28	221	29	216	29	224	28
12			230	28	232	27	231	28	226	28	234	28
13			239	27	242	27	241	27	236	28	243	27
14			248	26	250	26	250	27	245	27	252	26
15			257	26	259	26	259	26	254	27	261	26
16			265	26	267	25	268	26	263	26	269	26
17			274	25	275	25	276	25	271	26	277	25
18			281	25	283	25	284	25	280	26	285	25
19			289	25	290	24	292	25	288	25	293	25
20			297	25	298	24	300	25	296	25	301	25
21			304	24	305	24	307	25	304	25	308	24
22			312	24	313	24	315	25	312	25	316	24
23			319	24	320	24	322	24	319	25	323	24
24			327	24	327	24	330	24	327	25	331	24
25			334	24	334	24	337	24	335	25	338	24
26			342	24	342	24	345	24	342	25	345	24
27			349	24	349	24	352	24	350	25	353	24
28			357	25	356	24	360	25	358	25	360	25
29			364	25	364	24	368	25	366	25	368	25
30			372	25	371	25	375	25	374	25	376	25
31			380	25	379	25	383	25	382	26	384	25
32			388	25	387	25	391	25	390	26	392	25
33			397	26	395	25	399	26	399	26	400	26
34			405	26	403	26	408	26	407	26	408	26
35			414	27	411	26	417	27	416	27	417	27
36			423	27	420	27	426	27	426	27	426	27
37			432	28	429	27	435	28	435	28	436	28
38			442	28	439	28	445	28	445	29	446	28
39			453	29	449	29	455	29	456	29	456	29
40			464	30	460	30	466	30	467	30	467	30
41			476	31	472	31	478	31	479	32	479	31
42			489	33	484	32	491	33	492	33	492	33
43			503	34	498	34	505	34	507	35	506	34

44		518	36	514	36	520	36	523	37	522	37
45		536	39	531	39	538	39	541	39	540	39
46		557	43	552	43	560	43	562	43	562	43
47		584	49	579	49	586	49	589	49	588	49
48		600		600		600		600		600	
49		600		600		600		600		600	
50		600		600		600		600		600	

Table 2.5.21 RSSS Conversions for Grade 8 Plain English Mathematics

Raw Score	Paper		Online	
	Core 3		Core 3	
	SS	SEM	SS	SEM
0	0		0	
1	13	81	13	81
2	70	58	71	58
3	105	48	106	48
4	131	42	131	42
5	151	38	152	38
6	168	36	169	36
7	183	33	184	34
8	196	32	198	32
9	208	30	210	30
10	219	29	221	29
11	230	28	232	28
12	240	28	241	28
13	249	27	251	27
14	258	26	260	26
15	266	26	268	26
16	274	25	277	26
17	282	25	285	25
18	290	25	293	25
19	298	25	300	25
20	305	24	308	24
21	313	24	315	24
22	320	24	323	24
23	327	24	330	24
24	335	24	337	24
25	342	24	345	24
26	349	24	352	24
27	356	24	359	24
28	364	24	367	24
29	371	24	374	24
30	378	24	382	25
31	386	25	389	25
32	394	25	397	25
33	402	25	405	25
34	410	26	414	26
35	418	26	422	26
36	427	27	431	27
37	436	27	440	27
38	446	28	450	28
39	456	29	460	29
40	467	30	471	30
41	478	31	482	31
42	491	32	495	32
43	505	34	508	34

44	520	36	524	36
45	538	39	541	39
46	559	43	562	43
47	585	49	588	49
48	600		600	
49	600		600	
50	600		600	

Table 2.5.22 RSSS Conversions for Grade 8 Science

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0			0		0		0	
1			190	50	184	51	188	50
2			225	36	220	36	224	36
3			247	30	242	30	246	30
4			263	26	258	27	261	26
5			275	24	271	24	274	24
6			286	22	282	22	285	22
7			295	21	292	21	294	21
8			303	20	300	20	302	20
9			311	19	308	19	310	19
10			318	18	315	18	317	18
11			324	18	322	18	323	18
12			330	17	328	17	329	17
13			336	17	334	17	335	17
14			341	16	340	17	340	16
15			346	16	345	16	346	16
16			351	16	350	16	351	16
17			356	15	355	16	356	16
18			361	15	360	16	360	15
19			366	15	365	15	365	15
20			370	15	370	15	370	15
21			375	15	375	15	374	15
22			379	15	379	15	379	15
23			383	15	384	15	383	15
24			388	15	388	15	388	15
25			392	15	393	15	392	15
26			396	15	398	15	397	15
27			401	15	402	15	401	15
28			405	15	407	15	405	15
29			410	15	411	15	410	15
30			414	15	416	15	415	15
31			419	15	421	15	419	15
32			423	15	426	16	424	15
33			428	15	431	16	429	16
34			433	16	436	16	434	16
35			438	16	441	16	439	16
36			443	16	447	17	444	16
37			448	17	452	17	449	17
38			454	17	458	17	455	17
39			460	17	464	18	461	18
40			466	18	471	19	468	18
41			473	19	478	19	475	19
42			481	20	486	20	482	20
43			489	21	495	21	490	21

44		498	22	504	23	500	22
45		509	24	515	24	510	24
46		521	26	528	27	523	26
47		537	30	545	30	539	30
48		558	36	567	37	560	36
49		594	50	600		596	50
50		600		600		600	

Table 2.5.23 RSSS Conversions for United States History to 1877

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0		0		0	
1	157	55	157	55	145	55	144	55
2	196	40	197	40	185	40	184	40
3	220	33	221	33	209	33	209	33
4	238	29	238	29	226	29	227	29
5	252	26	252	26	240	26	241	27
6	264	25	264	25	253	25	253	25
7	274	23	275	23	263	23	264	23
8	284	22	284	22	273	22	274	22
9	292	21	293	21	281	21	283	21
10	300	20	301	20	289	21	291	21
11	308	20	308	20	297	20	299	20
12	315	19	316	19	304	19	306	20
13	322	19	322	19	311	19	313	19
14	328	19	329	19	318	19	320	19
15	335	18	335	18	324	18	326	19
16	341	18	342	18	330	18	333	19
17	347	18	348	18	336	18	339	18
18	353	18	354	18	342	18	345	18
19	359	18	360	18	348	18	351	18
20	365	18	366	18	354	18	357	18
21	370	18	372	18	360	18	364	18
22	376	18	378	18	366	18	370	18
23	382	18	384	18	372	18	376	18
24	388	18	390	18	378	18	382	19
25	395	18	396	19	385	18	389	19
26	401	19	403	19	391	19	395	19
27	408	19	409	19	398	19	402	19
28	414	19	416	19	405	19	409	20
29	421	20	423	20	412	20	416	20
30	429	20	431	21	419	21	424	21
31	437	21	439	21	427	21	432	21
32	446	22	448	22	436	22	441	22
33	455	23	457	23	445	23	451	23
34	466	25	468	25	456	25	462	25
35	477	26	480	26	468	26	474	27
36	492	29	494	29	482	29	488	29
37	509	33	512	33	500	33	506	33
38	533	40	536	40	524	40	531	40
39	573	55	575	55	563	55	570	55
40	600		600		600		600	

Table 2.5.24 RSSS Conversions for United States History 1877 to the Present

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0			0		0		0	
1			129	64	141	64	140	64
2			175	46	187	46	185	46
3			203	38	215	38	213	38
4			223	34	235	34	233	34
5			239	30	252	31	249	31
6			253	28	266	29	263	28
7			265	27	278	27	275	27
8			275	25	289	26	286	25
9			285	24	299	25	296	24
10			294	23	309	24	305	24
11			303	23	318	23	314	23
12			311	22	326	23	322	22
13			318	22	334	22	330	22
14			326	21	342	22	337	22
15			333	21	350	22	345	21
16			340	21	357	22	352	21
17			347	21	365	21	359	21
18			354	21	372	21	366	21
19			360	21	379	21	373	21
20			367	20	386	21	379	21
21			374	21	393	21	386	21
22			381	21	401	21	393	21
23			387	21	408	21	400	21
24			394	21	415	22	407	21
25			401	21	423	22	414	21
26			408	21	431	22	422	22
27			416	22	439	23	430	22
28			424	22	447	23	438	23
29			432	23	456	24	446	23
30			440	24	465	24	455	24
31			449	24	474	25	464	25
32			459	25	485	26	474	26
33			470	27	496	28	485	27
34			482	28	509	29	497	29
35			496	31	524	31	511	31
36			512	34	541	34	528	34
37			532	38	562	39	548	38
38			560	46	591	47	576	46
39			600		600		600	
40			600		600		600	

Table 2.5.25 RSSS Conversions for Civics and Economics

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0			0		0		0	
1			178	57	159	58	160	58
2			219	41	200	42	201	42
3			244	34	226	35	227	35
4			263	30	245	31	245	31
5			277	28	260	28	260	28
6			290	26	273	26	273	26
7			301	24	284	24	284	25
8			311	23	294	23	295	23
9			320	22	303	22	304	23
10			328	21	312	22	313	22
11			336	21	320	21	321	21
12			343	20	327	20	329	21
13			350	20	335	20	336	20
14			357	20	342	20	343	20
15			364	19	348	19	350	20
16			370	19	355	19	357	20
17			377	19	362	19	364	19
18			383	19	368	19	370	19
19			389	19	374	19	377	19
20			396	19	380	19	384	19
21			402	19	387	19	390	19
22			408	19	393	19	397	19
23			415	19	399	19	403	19
24			421	19	406	19	410	20
25			428	19	412	19	417	20
26			434	20	419	20	424	20
27			441	20	426	20	431	20
28			449	20	433	20	439	21
29			456	21	440	21	447	21
30			464	22	448	21	455	22
31			473	22	456	22	464	23
32			482	23	465	23	473	24
33			492	24	475	24	484	25
34			503	26	486	26	495	26
35			516	28	499	28	508	28
36			531	31	513	30	523	31
37			550	35	532	34	542	35
38			575	42	557	41	568	42
39			600		598	57	600	
40			600		600		600	

Table 2.5.26 RSSS Conversions for Virginia Studies

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0				0	
1	104	64	120	64			107	64
2	151	46	166	46			153	46
3	180	39	194	38			181	38
4	201	34	215	34			202	34
5	218	31	231	31			218	31
6	233	29	245	29			232	29
7	246	28	258	27			245	27
8	257	26	269	26			256	26
9	268	25	279	25			266	25
10	278	25	289	24			276	24
11	287	24	298	23			285	23
12	296	23	306	23			293	23
13	304	23	314	22			301	22
14	313	22	322	22			309	22
15	321	22	330	22			317	22
16	328	22	337	21			324	22
17	336	22	345	21			332	21
18	344	22	352	21			339	21
19	351	22	359	21			346	21
20	358	22	366	21			353	21
21	366	22	373	21			360	21
22	373	22	380	21			368	21
23	381	22	388	21			375	21
24	389	22	395	21			382	21
25	396	22	402	22			389	22
26	404	22	410	22			397	22
27	412	23	418	22			405	22
28	421	23	426	23			413	23
29	430	24	434	23			422	23
30	439	24	443	24			430	24
31	449	25	453	25			440	25
32	460	26	463	26			450	26
33	471	28	474	27			461	27
34	484	29	486	29			474	29
35	499	31	500	31			488	31
36	516	34	517	34			504	34
37	537	39	537	38			525	38
38	566	46	565	46			552	46
39	600		600				598	64
40	600		600				600	

Table 2.5.27 RSSS Conversions for EOC Reading

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							109	66	90	66	100	66
2							156	47	138	48	147	48
3							184	39	168	40	176	39
4							205	35	189	35	197	35
5							222	31	206	32	214	32
6							236	29	221	30	228	29
7							248	27	234	28	240	28
8							259	26	245	27	252	26
9							269	25	256	26	262	25
10							278	24	266	25	271	24
11							287	23	275	24	280	24
12							295	23	283	23	288	23
13							303	22	291	23	296	22
14							310	22	299	22	304	22
15							317	21	307	22	311	21
16							324	21	314	22	318	21
17							331	21	321	21	325	21
18							337	20	328	21	331	21
19							343	20	335	21	338	20
20							350	20	342	21	344	20
21							356	20	348	21	351	20
22							362	20	355	21	357	20
23							368	20	361	20	363	20
24							374	20	368	20	369	20
25							380	20	374	20	375	20
26							386	20	381	20	381	20
27							392	20	387	21	388	20
28							398	20	394	21	394	20
29							404	20	400	21	400	20
30							410	20	407	21	406	20
31							416	20	414	21	413	20
32							423	20	420	21	419	21
33							429	21	427	21	426	21
34							436	21	435	22	433	21
35							443	21	442	22	440	22
36							450	22	450	22	447	22
37							458	22	458	23	455	22
38							466	23	466	23	462	23
39							474	23	475	24	471	24
40							483	24	484	25	480	24
41							492	25	494	26	489	25
42							502	26	504	27	499	26
43							513	28	516	28	510	28

44		526	29	529	30	523	29
45		540	32	543	32	537	32
46		557	35	561	35	554	35
47		579	40	583	40	575	39
48		600		600		600	
49		600		600		600	
50		600		600		600	

Table 2.5.28 RSSS Conversions for EOC Earth Science

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							165	52	162	52	167	52
2							202	37	199	37	204	37
3							225	31	222	31	227	31
4							241	27	238	27	243	27
5							254	25	252	25	256	25
6							265	23	263	23	268	23
7							275	22	273	22	277	22
8							283	20	281	21	286	21
9							291	20	289	20	294	20
10							298	19	296	19	301	19
11							305	18	303	18	308	18
12							311	18	310	18	314	18
13							317	17	316	17	320	17
14							323	17	322	17	326	17
15							329	17	327	17	332	17
16							334	16	333	17	337	17
17							339	16	338	16	343	16
18							344	16	343	16	348	16
19							349	16	348	16	353	16
20							354	16	353	16	358	16
21							359	16	358	16	362	16
22							363	16	362	16	367	16
23							368	15	367	16	372	16
24							373	15	372	16	377	16
25							377	15	377	15	381	16
26							382	15	381	16	386	16
27							387	16	386	16	391	16
28							391	16	391	16	396	16
29							396	16	395	16	400	16
30							401	16	400	16	405	16
31							406	16	405	16	410	16
32							411	16	410	16	415	16
33							416	16	415	16	420	16
34							421	17	420	17	425	17
35							427	17	426	17	431	17
36							432	17	431	17	436	17
37							438	17	437	17	442	17
38							444	18	443	18	448	18
39							451	18	450	18	455	18
40							458	19	457	19	461	19
41							465	20	464	20	469	20
42							473	21	472	21	477	21
43							482	22	481	22	485	22

44		491	23	490	23	495	23
45		503	25	502	25	506	25
46		516	27	515	27	519	27
47		533	31	531	31	536	31
48		555	38	554	38	559	37
49		593	52	591	52	596	52
50		600		600		600	

Table 2.5.29 RSSS Conversions for EOC Biology

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							216	44	221	44	214	44
2							248	32	253	32	245	32
3							267	26	272	26	265	26
4							281	23	285	23	279	23
5							292	21	297	21	290	21
6							302	20	306	19	299	20
7							310	18	314	18	308	18
8							318	17	321	17	315	17
9							324	17	328	16	322	17
10							330	16	334	16	328	16
11							336	16	339	15	334	15
12							342	15	345	15	339	15
13							347	15	350	15	344	15
14							352	14	354	14	349	14
15							356	14	359	14	354	14
16							361	14	363	14	358	14
17							365	14	368	13	362	14
18							370	14	372	13	367	13
19							374	13	376	13	371	13
20							378	13	380	13	375	13
21							382	13	384	13	379	13
22							386	13	387	13	383	13
23							390	13	391	13	387	13
24							394	13	395	13	391	13
25							398	13	399	13	394	13
26							401	13	403	13	398	13
27							405	13	406	13	402	13
28							409	13	410	13	406	13
29							413	13	414	13	410	13
30							417	13	418	13	414	13
31							421	13	422	13	418	13
32							425	13	426	13	422	13
33							430	14	430	13	426	14
34							434	14	434	14	430	14
35							438	14	438	14	435	14
36							443	14	443	14	439	14
37							448	15	448	14	444	15
38							453	15	452	15	449	15
39							458	15	458	15	454	15
40							464	16	463	16	460	16
41							470	16	469	16	466	17
42							476	17	475	17	473	17
43							483	18	483	18	480	18

44		491	19	490	19	488	19
45		500	21	500	21	497	21
46		511	23	510	23	509	23
47		525	26	524	26	522	26
48		544	31	543	31	542	32
49		575	44	574	44	573	44
50		600		600		600	

Table 2.5.30 RSSS Conversions for EOC Chemistry

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							199	46	203	45	204	45
2							232	33	235	33	237	33
3							252	27	255	27	257	27
4							267	24	270	24	271	24
5							279	22	281	22	283	22
6							290	20	291	20	293	20
7							298	19	300	19	302	19
8							306	18	308	18	310	18
9							314	18	315	17	317	17
10							320	17	322	17	323	17
11							327	16	328	16	329	16
12							333	16	334	16	335	16
13							338	16	339	15	341	15
14							343	15	344	15	346	15
15							349	15	349	15	351	15
16							353	15	354	15	356	15
17							358	14	359	14	361	14
18							363	14	364	14	365	14
19							368	14	368	14	370	14
20							372	14	373	14	374	14
21							376	14	377	14	378	14
22							381	14	382	14	383	14
23							385	14	386	14	387	14
24							389	14	390	14	391	14
25							394	14	394	14	395	14
26							398	14	399	14	400	14
27							402	14	403	14	404	14
28							407	14	407	14	408	14
29							411	14	412	14	412	14
30							415	14	416	14	417	14
31							420	14	421	14	421	14
32							424	14	425	14	425	14
33							429	14	430	14	430	14
34							434	15	434	15	435	14
35							439	15	439	15	439	15
36							444	15	444	15	444	15
37							449	15	449	15	449	15
38							455	16	455	16	455	16
39							460	16	461	16	460	16
40							467	17	467	17	466	16
41							473	17	473	17	473	17
42							480	18	480	18	479	18
43							488	19	488	19	487	19

44		497	20	496	20	495	20
45		507	22	506	22	505	21
46		519	24	518	24	516	24
47		533	27	532	27	531	27
48		553	33	552	32	550	32
49		586	45	584	45	582	45
50		600		600		600	

Table 2.5.31 RSSS Conversions for EOC Algebra I

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							207	46	219	45	210	45
2							240	33	251	32	243	33
3							260	28	271	27	263	27
4							275	24	285	24	277	24
5							287	22	296	22	289	22
6							297	21	306	20	299	20
7							306	19	314	19	307	19
8							314	18	322	18	315	18
9							322	18	329	17	322	17
10							328	17	335	16	329	17
11							335	16	341	16	335	16
12							341	16	347	15	340	16
13							346	16	352	15	346	15
14							352	15	357	15	351	15
15							357	15	362	15	356	15
16							362	15	366	14	361	14
17							367	15	371	14	365	14
18							371	14	375	14	370	14
19							376	14	380	14	374	14
20							380	14	384	14	378	14
21							385	14	388	14	383	14
22							389	14	392	14	387	14
23							393	14	396	13	391	14
24							398	14	400	13	395	14
25							402	14	404	13	400	14
26							406	14	408	13	404	14
27							410	14	413	13	408	14
28							415	14	417	14	412	14
29							419	14	421	14	416	14
30							423	14	425	14	421	14
31							428	14	429	14	425	14
32							432	14	434	14	429	14
33							437	14	438	14	434	14
34							442	15	443	14	439	15
35							446	15	447	15	443	15
36							451	15	452	15	448	15
37							457	15	457	15	454	15
38							462	16	462	15	459	16
39							468	16	468	16	465	16
40							474	17	474	16	471	17
41							480	17	480	17	477	17
42							488	18	487	18	484	18
43							495	19	494	19	492	19

44		504	20	503	20	501	20
45		514	22	512	21	511	22
46		526	24	524	24	522	24
47		540	27	538	27	537	27
48		560	33	558	32	557	33
49		593	45	590	45	589	45
50		600		600		600	

Table 2.5.32 RSSS Conversions for EOC Plain English Algebra I

Raw Score	Paper		Online	
	Core 3		Core 3	
	SS	SEM	SS	SEM
0			0	
1			211	45
2			244	33
3			264	27
4			278	24
5			290	22
6			300	20
7			308	19
8			316	18
9			323	17
10			330	17
11			336	16
12			341	16
13			347	15
14			352	15
15			357	15
16			362	14
17			366	14
18			371	14
19			375	14
20			380	14
21			384	14
22			388	14
23			392	14
24			396	14
25			401	14
26			405	14
27			409	14
28			413	14
29			417	14
30			422	14
31			426	14
32			430	14
33			435	14
34			439	14
35			444	15
36			449	15
37			454	15
38			459	16
39			465	16
40			471	17
41			477	17
42			484	18
43			492	19

44		500	20
45		510	22
46		522	24
47		536	27
48		555	33
49		588	45
50		600	

Table 2.5.33 RSSS Conversions for EOC Geometry

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							175	49	181	49	173	50
2							210	35	217	35	209	36
3							231	29	238	29	231	30
4							247	26	254	26	248	27
5							260	24	266	24	261	24
6							271	22	277	22	272	22
7							280	21	286	21	282	21
8							289	20	294	20	291	20
9							296	19	302	19	299	19
10							304	18	309	18	307	19
11							310	18	315	18	314	18
12							317	17	322	17	320	18
13							323	17	327	17	326	17
14							329	17	333	16	332	17
15							334	16	339	16	338	17
16							340	16	344	16	344	16
17							345	16	349	16	349	16
18							351	16	354	16	355	16
19							356	16	359	15	360	16
20							361	16	364	15	365	16
21							366	16	369	15	370	16
22							371	16	374	15	375	16
23							376	16	378	15	380	16
24							381	16	383	15	385	16
25							387	16	388	15	391	16
26							392	16	393	15	396	16
27							397	16	398	16	401	16
28							403	16	403	16	406	16
29							408	16	408	16	411	16
30							414	17	413	16	417	16
31							420	17	419	16	422	17
32							426	17	424	17	428	17
33							432	18	430	17	434	17
34							439	18	436	17	441	18
35							446	19	443	18	447	18
36							454	19	450	19	454	19
37							462	20	457	19	462	20
38							471	21	465	20	470	21
39							481	23	475	22	480	22
40							492	24	485	23	490	24
41							505	27	498	26	503	26
42							522	30	513	29	519	29
43							544	36	534	35	540	35

44		580	50	569	49	575	49
45		600		600		600	

Table 2.5.34 RSSS Conversions for EOC Algebra II

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							168	52	167	51	158	51
2							205	37	203	37	195	37
3							227	31	225	30	217	31
4							244	27	241	27	233	27
5							257	25	254	24	246	24
6							268	23	264	22	257	23
7							277	21	274	21	266	21
8							286	20	282	20	275	20
9							293	19	289	19	282	19
10							301	19	296	18	289	18
11							307	18	303	18	296	18
12							313	18	309	17	302	17
13							319	17	314	17	308	17
14							325	17	320	16	313	17
15							330	16	325	16	318	16
16							336	16	330	16	324	16
17							341	16	335	16	329	16
18							346	16	340	15	333	16
19							350	16	344	15	338	15
20							355	15	349	15	343	15
21							360	15	353	15	347	15
22							364	15	358	15	352	15
23							369	15	362	15	356	15
24							373	15	366	15	361	15
25							378	15	371	15	365	15
26							383	15	375	15	370	15
27							387	15	380	15	374	15
28							392	15	384	15	378	15
29							396	15	388	15	383	15
30							401	15	393	15	387	15
31							405	16	397	15	392	15
32							410	16	402	15	397	16
33							415	16	407	16	402	16
34							420	16	412	16	407	16
35							425	16	417	16	412	16
36							431	17	422	16	417	17
37							436	17	427	17	422	17
38							442	17	433	17	428	17
39							448	18	439	18	434	18
40							455	19	446	18	441	18
41							462	19	453	19	448	19
42							470	20	460	20	455	20
43							478	21	468	21	464	21

44		488	23	478	22	473	22
45		498	24	488	24	484	24
46		511	27	501	27	497	27
47		527	31	517	30	513	30
48		550	37	539	37	535	37
49		586	51	576	51	571	51
50		600		600		600	

Table 2.5.35 RSSS Conversions for EOC Virginia and US History

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							177	52	174	52	176	52
2							214	37	211	37	213	37
3							236	31	233	31	235	31
4							253	27	249	27	252	27
5							266	25	262	25	265	25
6							277	23	273	23	276	23
7							287	22	283	21	285	21
8							295	20	291	20	294	20
9							303	19	298	19	301	19
10							310	19	305	18	308	19
11							317	18	312	18	315	18
12							323	18	318	17	321	17
13							329	17	323	17	327	17
14							334	17	329	16	332	17
15							340	16	334	16	337	16
16							345	16	339	16	342	16
17							350	16	343	15	347	16
18							354	15	348	15	352	15
19							359	15	352	15	356	15
20							364	15	357	15	361	15
21							368	15	361	15	365	15
22							372	15	365	14	369	15
23							376	15	369	14	374	15
24							380	15	373	14	378	15
25							385	14	377	14	382	14
26							389	14	381	14	386	14
27							393	14	385	14	390	14
28							397	14	389	14	394	14
29							400	14	392	14	398	14
30							404	14	396	14	402	14
31							408	14	400	14	406	14
32							412	14	404	14	410	14
33							416	14	408	14	414	14
34							420	14	412	14	418	14
35							424	14	416	14	422	15
36							428	14	420	14	426	15
37							432	15	424	14	430	15
38							436	15	428	15	434	15
39							441	15	432	15	439	15
40							445	15	436	15	443	15
41							449	15	441	15	448	15
42							454	15	445	15	452	16
43							458	16	450	16	457	16

44		463	16	455	16	462	16
45		468	16	460	16	467	16
46		473	16	465	17	473	17
47		479	17	470	17	478	17
48		484	17	476	17	484	18
49		490	18	482	18	491	18
50		497	18	489	19	497	19
51		504	19	496	19	505	20
52		511	20	503	20	513	21
53		520	21	512	21	521	22
54		529	23	521	23	531	23
55		540	24	532	25	543	25
56		553	27	545	27	556	28
57		569	31	562	31	573	31
58		591	37	584	37	596	38
59		600		600		600	
60		600		600		600	

Table 2.5.36 RSSS Conversions for EOC World History I

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							170	57	166	57	163	57
2							210	40	206	41	204	41
3							234	33	230	33	228	34
4							251	29	248	29	246	30
5							265	27	261	27	260	27
6							277	25	273	25	272	25
7							287	23	283	23	282	23
8							296	22	292	22	291	22
9							304	21	300	21	300	21
10							311	20	308	20	307	20
11							318	19	315	19	315	20
12							325	19	321	19	321	19
13							331	18	327	18	327	19
14							336	18	333	18	333	18
15							342	17	338	17	339	18
16							347	17	344	17	345	17
17							352	17	349	17	350	17
18							357	16	354	16	355	17
19							362	16	358	16	360	17
20							367	16	363	16	365	16
21							371	16	368	16	370	16
22							376	16	372	16	374	16
23							380	16	376	16	379	16
24							384	15	381	15	384	16
25							389	15	385	15	388	16
26							393	15	389	15	392	16
27							397	15	393	15	397	16
28							401	15	397	15	401	16
29							405	15	402	15	406	16
30							409	15	406	15	410	16
31							414	15	410	15	414	16
32							418	15	414	15	418	16
33							422	15	418	15	423	16
34							426	15	422	15	427	16
35							430	15	426	15	432	16
36							435	16	431	15	436	16
37							439	16	435	16	441	16
38							443	16	439	16	445	16
39							448	16	444	16	450	16
40							453	16	448	16	455	16
41							457	16	453	16	459	17
42							462	17	458	16	464	17
43							467	17	463	17	469	17

44		472	17	468	17	475	17
45		478	17	473	17	480	18
46		483	18	478	18	486	18
47		489	18	484	18	492	18
48		495	19	490	19	498	19
49		502	19	497	19	505	20
50		509	20	503	20	512	20
51		516	21	511	21	519	21
52		524	22	519	22	527	22
53		534	23	528	23	537	23
54		544	25	538	25	547	25
55		556	27	550	27	559	27
56		570	29	563	29	573	29
57		587	34	581	33	590	34
58		600		600		600	
59		600		600		600	
60		600		600		600	

Table 2.5.37 RSSS Conversions for EOC World History II

Raw Score	Paper						Online					
	Core 1		Core 2		Core 3		Core 1		Core 2		Core 3	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0							0		0		0	
1							163	56	156	56	146	56
2							203	40	196	40	186	40
3							227	33	220	33	211	34
4							245	29	237	29	228	30
5							259	27	251	26	243	27
6							270	25	263	24	255	25
7							281	23	273	23	265	23
8							290	22	282	22	275	22
9							298	21	290	21	283	21
10							305	20	298	20	291	20
11							312	19	305	19	298	20
12							319	19	311	19	305	19
13							325	18	317	18	311	19
14							331	18	323	18	317	18
15							337	17	328	17	323	18
16							342	17	334	17	329	17
17							347	17	339	17	334	17
18							352	16	344	16	339	17
19							357	16	348	16	344	17
20							362	16	353	16	349	16
21							366	16	358	16	354	16
22							371	16	362	16	359	16
23							375	16	366	15	363	16
24							379	15	371	15	368	16
25							384	15	375	15	372	16
26							388	15	379	15	377	16
27							392	15	383	15	381	16
28							396	15	388	15	386	16
29							400	15	392	15	390	16
30							404	15	396	15	394	16
31							409	15	400	15	399	16
32							413	15	404	15	403	16
33							417	15	408	15	407	16
34							421	15	412	15	412	16
35							425	15	416	15	416	16
36							429	15	421	15	421	16
37							434	15	425	15	425	16
38							438	16	429	16	430	16
39							443	16	434	16	434	16
40							447	16	438	16	439	16
41							452	16	443	16	444	16
42							456	16	448	16	449	17
43							461	17	452	17	454	17

44		466	17	457	17	459	17
45		472	17	463	17	465	18
46		477	18	468	18	471	18
47		483	18	474	18	477	18
48		489	18	480	18	483	19
49		495	19	486	19	490	19
50		502	20	493	20	497	20
51		509	21	500	21	504	21
52		517	22	509	22	513	22
53		526	23	517	23	522	23
54		536	24	527	24	532	25
55		548	26	539	26	544	27
56		561	29	553	29	558	29
57		579	33	570	33	576	33
58		600		594	40	600	
59		600		600		600	
60		600		600		600	

Table 2.5.38 RSSS Conversions for EOC World Geography

Raw Score	Paper				Online			
	Core 1		Core 2		Core 1		Core 2	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0					0		0	
1					101	67	88	67
2					149	48	135	48
3					177	40	164	40
4					198	35	185	35
5					215	32	201	32
6					229	29	215	29
7					241	27	227	27
8					251	26	238	26
9					261	25	248	25
10					270	24	257	24
11					278	23	265	23
12					286	22	273	22
13					293	22	280	22
14					300	21	287	21
15					307	21	294	21
16					313	20	300	20
17					320	20	307	20
18					326	20	313	20
19					331	19	318	19
20					337	19	324	19
21					343	19	330	19
22					348	19	335	19
23					353	19	340	19
24					358	18	346	19
25					364	18	351	18
26					369	18	356	18
27					374	18	361	18
28					379	18	366	18
29					384	18	371	18
30					389	18	376	18
31					394	18	381	18
32					399	18	386	18
33					404	18	391	18
34					409	18	397	18
35					414	18	402	18
36					419	18	407	19
37					424	19	412	19
38					429	19	418	19
39					435	19	423	19
40					440	19	429	19
41					446	19	434	19
42					452	20	440	20
43					457	20	446	20

44		464	20	453	20
45		470	21	459	21
46		476	21	466	21
47		483	22	473	22
48		490	22	480	22
49		498	23	488	23
50		506	24	496	24
51		515	25	505	25
52		525	26	515	26
53		535	27	526	27
54		547	29	538	29
55		561	31	552	31
56		577	35	568	35
57		598	39	589	40
58		600		600	
59		600		600	
60		600		600	

Table 2.5.39 RSSS Conversions for Grade 5 Writing

Raw Score	Core 1				Core 2				Core 3			
	Prompt 5255		Prompt 5264		Prompt 5255		Prompt 5264		Prompt 5255		Prompt 5264	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0		0		0		0		0	
1	21		21		22		22		22		22	
2	42		42		44		44		45		44	
3	63		63		66		66		67		67	
4	84		83		88		88		89		89	
5	105		104		110		110		112		111	
6	126	94	125	94	132	94	132	94	134	94	133	94
7	192	48	192	48	197	47	197	47	199	47	199	47
8	226	34	226	34	230	33	231	34	231	33	232	34
9	246	28	247	28	249	27	251	28	250	27	252	28
10	260	25	262	25	263	24	265	24	264	24	267	24
11	272	22	274	23	274	22	277	22	276	22	278	22
12	282	21	285	21	284	20	286	21	285	20	288	21
13	291	20	294	20	292	19	295	20	294	19	297	20
14	299	19	302	19	300	19	303	19	301	19	305	19
15	306	19	310	19	307	18	310	18	309	18	312	18
16	313	18	317	18	314	18	317	18	316	18	319	18
17	321	18	324	18	321	18	324	18	323	18	326	18
18	328	18	331	18	328	18	331	18	330	18	333	18
19	335	18	338	18	335	18	337	18	337	18	339	18
20	342	18	345	18	341	18	344	18	344	18	346	18
21	349	19	352	18	348	18	351	18	351	18	353	18
22	357	19	359	18	355	18	357	18	358	18	360	18
23	364	19	366	18	363	19	364	18	365	19	367	18
24	372	19	373	19	370	19	371	18	372	19	374	18
25	380	19	381	19	378	19	378	18	380	19	381	18
26	388	19	388	19	385	19	386	18	387	19	388	18
27	396	20	396	19	393	19	393	19	395	19	395	18
28	405	20	404	19	401	19	400	19	403	19	402	19
29	413	20	411	19	409	20	408	19	411	19	410	19
30	421	20	419	19	417	20	416	19	419	20	417	19
31	430	20	428	20	426	20	424	19	427	20	425	19
32	439	21	436	20	434	20	432	20	436	20	433	20
33	448	21	445	21	443	21	440	20	444	21	441	20
34	458	22	454	21	453	21	449	21	454	21	450	21
35	468	22	464	22	463	22	459	22	464	22	460	22
36	479	24	475	23	474	23	470	23	475	23	471	23
37	492	25	487	25	486	25	482	25	487	25	482	24
38	506	27	501	26	501	27	496	26	501	27	496	26
39	523	29	517	28	518	29	512	28	518	29	512	28
40	542	31	535	30	538	32	530	30	538	31	531	30
41	563	33	555	32	560	33	551	33	560	33	552	32
42	588	36	579	36	585	36	576	36	585	36	576	36
43	600		600		600		600		600		600	

44	600		600		600		600		600		600	

Table 2.5.40 RSSS Conversions for Grade 8 Writing

Raw Score	Core 1				Core 2				Core 3			
	Prompt 8527		Prompt 8529		Prompt 8527		Prompt 8529		Prompt 8527		Prompt 8529	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0		0		0		0		0	
1	44		44		43		44		44		44	
2	87		87		87		87		87		87	
3	131		131		130		131		131		131	
4	174		175		173		174		175		175	
5	218		218		217		218		218		218	
6	261	48	262	48	260	48	261	48	262	48	262	48
7	295	24	296	24	294	24	295	24	295	24	296	24
8	312	17	313	17	311	17	312	17	313	17	313	17
9	322	14	323	14	322	14	322	14	323	14	323	14
10	330	13	330	12	329	13	330	12	330	12	331	12
11	336	11	336	11	335	11	335	11	336	11	336	11
12	341	11	341	10	340	11	340	10	341	10	341	10
13	345	10	345	10	344	10	345	10	345	10	346	10
14	349	10	349	9	348	9	348	9	349	9	349	9
15	353	9	353	9	352	9	352	9	353	9	353	9
16	356	9	356	9	355	9	355	9	356	9	356	9
17	359	9	360	9	359	9	359	9	359	9	360	9
18	363	9	363	9	362	9	362	9	363	9	363	9
19	366	9	366	9	365	9	365	9	366	9	366	9
20	369	9	369	9	368	9	368	9	369	8	369	9
21	372	9	373	9	371	9	372	9	372	8	372	9
22	375	9	376	9	374	9	375	9	375	8	375	9
23	378	9	379	9	377	9	378	9	378	8	378	9
24	381	9	382	9	380	9	381	9	380	8	381	9
25	384	9	386	9	383	9	384	9	383	8	384	9
26	387	9	389	9	386	9	388	9	386	8	388	9
27	390	9	392	9	389	9	391	9	389	8	391	9
28	394	9	395	9	393	9	394	9	392	8	394	9
29	397	9	398	9	396	9	397	9	395	8	397	9
30	400	9	402	9	399	9	401	9	398	9	400	9
31	403	9	405	9	402	9	404	9	401	9	403	9
32	406	9	409	9	406	9	408	9	404	9	406	9
33	410	9	412	9	409	9	411	9	407	9	410	9
34	413	9	416	10	413	10	415	10	411	9	413	9
35	417	10	420	10	417	10	419	10	414	9	417	10
36	421	10	424	10	421	10	423	10	418	10	421	10
37	425	10	428	10	425	10	428	11	422	10	425	10
38	430	10	433	11	430	11	433	11	426	10	429	11
39	435	11	438	11	434	11	438	11	431	11	434	11
40	440	11	444	12	440	11	444	12	436	11	440	12
41	445	12	450	12	445	12	450	13	441	12	446	13
42	451	12	457	13	451	12	457	13	447	12	453	13
43	457	13	464	14	457	13	464	14	454	13	461	14

44	464	13	472	14	464	13	473	14	461	14	469	15
45	472	15	482	16	473	15	482	16	469	15	479	16
46	483	17	493	18	483	17	493	18	480	17	491	18
47	500	24	510	24	500	24	510	24	497	24	508	24
48	600		600		600		600		600		600	

Table 2.5.41 RSSS Conversions for EOC Writing

Raw Score	Core 1				Core 2				Core 3			
	Prompt 1656		Prompt 1663		Prompt 1656		Prompt 1663		Prompt 1656		Prompt 1663	
	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM	SS	SEM
0	0		0		0		0		0		0	
1	21		20		22		21		23		22	
2	41		41		43		42		45		44	
3	62		61		65		64		68		67	
4	83		81		86		85		90		89	
5	103		102		108		106		113		111	
6	124	89	122	89	129	89	127	89	135	89	133	89
7	186	45	185	45	192	45	190	45	197	45	195	45
8	218	32	216	32	223	32	221	32	228	32	226	32
9	237	26	235	26	242	26	240	26	247	26	244	26
10	250	23	249	23	255	23	253	23	260	23	258	23
11	261	21	259	21	266	21	264	21	270	20	268	21
12	270	19	268	19	274	19	273	19	279	19	277	19
13	278	18	276	18	282	18	281	18	286	18	285	18
14	285	17	283	17	289	17	288	17	293	17	292	17
15	291	17	290	17	295	17	294	17	299	17	298	17
16	297	16	296	16	301	16	300	16	305	16	305	16
17	302	16	302	16	307	16	306	16	311	16	310	16
18	308	15	307	16	313	16	312	16	317	16	316	16
19	313	15	312	15	318	15	317	16	322	15	321	16
20	318	15	318	15	323	15	323	15	327	15	327	15
21	323	15	323	15	328	15	328	15	333	15	332	15
22	328	15	328	15	334	15	333	15	338	15	337	15
23	333	15	333	15	339	15	338	15	343	15	342	15
24	338	15	338	15	344	15	343	15	348	15	347	15
25	343	15	343	15	349	15	349	15	353	15	353	15
26	349	15	348	15	354	15	354	15	358	15	358	15
27	354	15	353	15	360	15	359	15	364	15	363	15
28	359	15	358	15	365	15	364	15	369	15	368	15
29	364	15	363	15	370	15	369	15	374	15	373	15
30	370	16	368	15	375	15	374	15	379	15	378	15
31	375	16	374	15	381	15	379	15	384	15	383	15
32	380	16	379	15	386	15	384	15	389	15	388	15
33	386	16	384	15	391	16	390	15	395	15	393	15
34	392	16	390	16	397	16	395	15	400	15	398	15
35	397	16	395	16	402	16	400	16	405	15	403	15
36	403	16	401	16	408	16	406	16	411	16	408	16
37	409	16	407	16	414	16	411	16	416	16	414	16
38	415	17	412	16	419	16	417	16	422	16	419	16
39	421	17	419	17	426	17	423	17	428	17	425	17
40	428	17	425	17	432	17	429	17	434	17	432	17
41	435	18	432	18	439	18	436	18	441	18	438	18
42	442	19	440	19	446	19	444	19	448	18	446	18
43	450	20	448	20	454	19	452	19	456	19	454	19

44	460	21	457	21	463	21	461	21	465	21	463	21
45	470	23	467	22	474	22	471	22	475	22	473	22
46	483	24	479	24	486	24	482	24	487	24	484	23
47	497	26	493	26	500	26	496	25	501	26	497	25
48	513	28	509	27	515	27	511	26	516	27	512	26
49	531	29	525	28	533	28	527	28	534	28	528	27
50	550	29	543	29	551	29	545	29	552	29	545	28
51	570	30	563	31	570	30	564	30	571	30	564	30
52	592	33	586	34	593	33	587	34	593	33	587	34
53	600		600		600		600		600		600	
54	600		600		600		600		600		600	

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, D.C.: American Psychological Association.

Andrich, A., & Luo, G. (2004). *Modern measurement and analysis in social science*. Perth, Western Australia: Murdoch University.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 508-600). Washington, D. C.: American Council on Education.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Harcourt Brace Jovanovich College Publishers.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York: McGraw-Hill.

Haertel, E. H. (1996). *Estimating the decision consistency from a single administration of a performance assessment battery. A report on the National Board of Professional Teaching Standards McGEN Assessment.* Palo Alto, CA: Stanford University.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*, 139–164.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kerlinger, F. N. (1973). *Foundations of behavioral research.* New York: Holt, Rinehart, & Winston.

Kolen, M. J., & Brennan, R. L. (2005). *Test equating, scaling, and linking: Methods and practices* (2nd Ed.). New York: Springer-Verlag.

Linacre, J. M. (2006). *WINSTEPS: Rasch measurement computer programs*. Chicago: Winsteps.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement, 32*, 179-197.

Loomis, S.C. & Bourque, M. L. (2001). From tradition to innovation : Standard setting on the National Assessment of Educational Progress. In G. J. Cizek (Ed.), *Setting performance standards* (pp. 175-218). Mahwah, NJ: Lawrence Erlbaum Associates.

Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score "equatings." *Applied Psychological Measurement, 8*, 452-461.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards* (pp. 249-282). Mahwah, NJ: Lawrence Erlbaum Associates.

Nunnaly, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: University of Chicago Press.

Reckase, M. D. (2000, June). *The evolution of the NAEP Achievement Levels Setting Process: A summary of the research and development efforts conducted by ACT.* (A report for the National Assessment Governing Board, Washington, DC). Iowa City, IA: ACT, Inc.

Suen, H. K. (1990). *Principles of test theories*. Hillsdale, NJ: Lawrence.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika, 51*, 567-577.

Wright, B. D., & Stone, M. H. (1979). *Best test design.* Chicago, IL: MESA Press.

Young, M. J., & Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment.* (CSE Technical Report 475). Center for the Study of Evaluation, Standards, and Student Testing. Los Angeles, CA: University of California, Los Angeles.